

Using D2K Data Mining Platform for Understanding the Dynamic Evolution of Land-Surface Variables



Praveen Kumar¹, Peter Bajcsy², David Tcheng²,
David Clutter², Vikas Mehra¹, Wei-Wen Feng²,
Pratyush Sinha¹, and Amanda B. White¹

¹Civil and Environmental Engineering

²National Center For Supercomputing Applications
University of Illinois, Urbana-Champaign



Praveen Kumar
Assoc. Prof.
Civil & Env. Eng.

Peter Bajcsy
Res. Sci. NCSA



David Tcheng
NCSA



David Clutter
NCSA

UIUC Team



(Not Shown)
Vikas Mehra
Grad Student
Civil & Env. Eng.



Pratyush Sinha
Grad Student
Civil and Env. Eng.



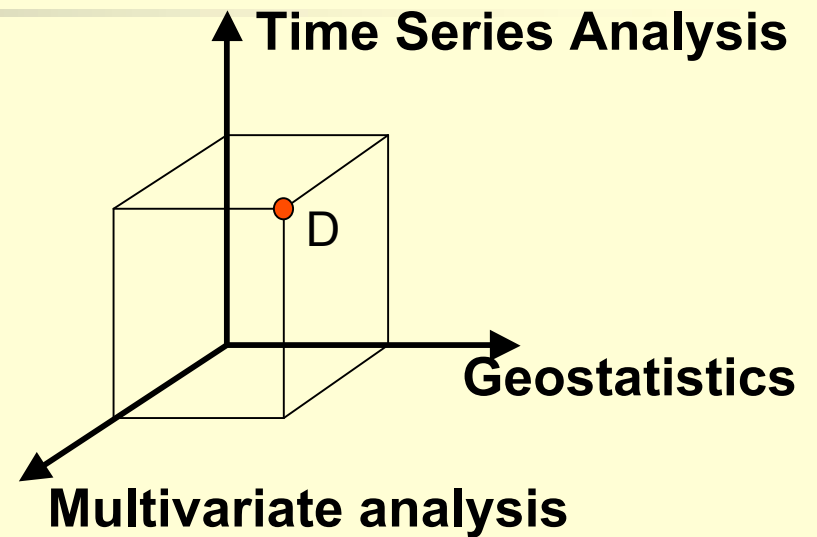
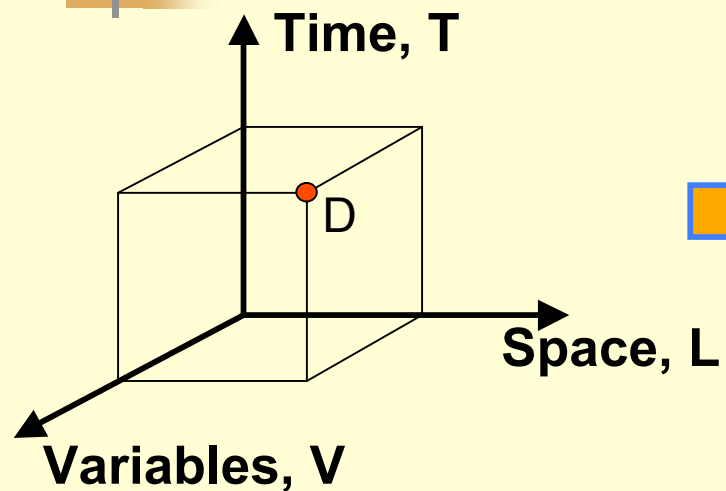
Wei-Wen Feng
Grad Student
Comp. Sci.



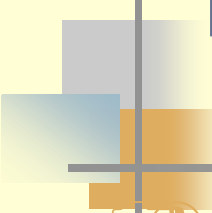
Amanda White
Grad Student
Civil & Env. Eng.





Analyses Needs



Research and Development Focus



 Develop KDD (Knowledge Discovery in Databases) techniques to exploit geo-referenced, spatio-temporal datasets including satellite data

 Use these techniques to obtain a better understanding of the relationships between vegetation amount and vigor and topography, climate, physiography, and hydrology

KDD (Data Mining) is an emerging field for scientific discovery → find relationships in large, complex databases

Topographic Influence on Vegetation



Large Scale Topographic Influence



Small Scale Soil, Land Cover, Topographic Influence



Influence of Slope → Water Availability

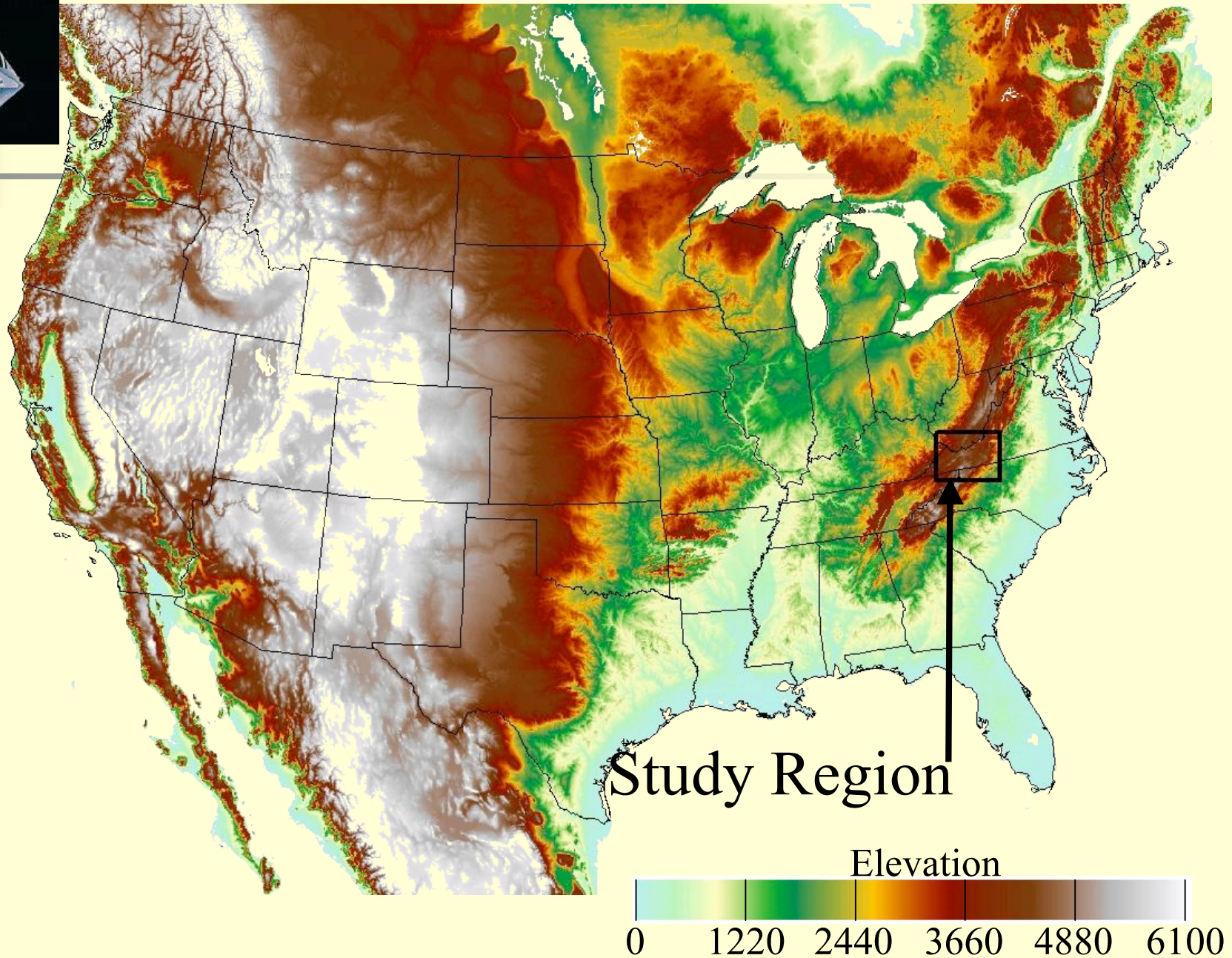
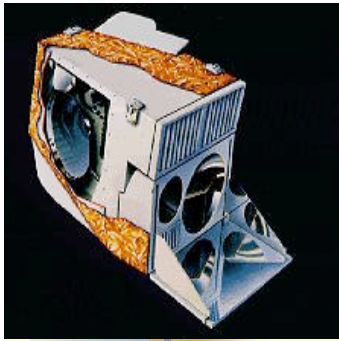


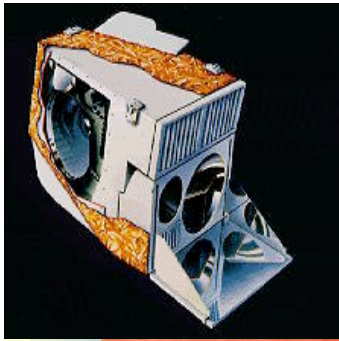
Top of King's Canyon



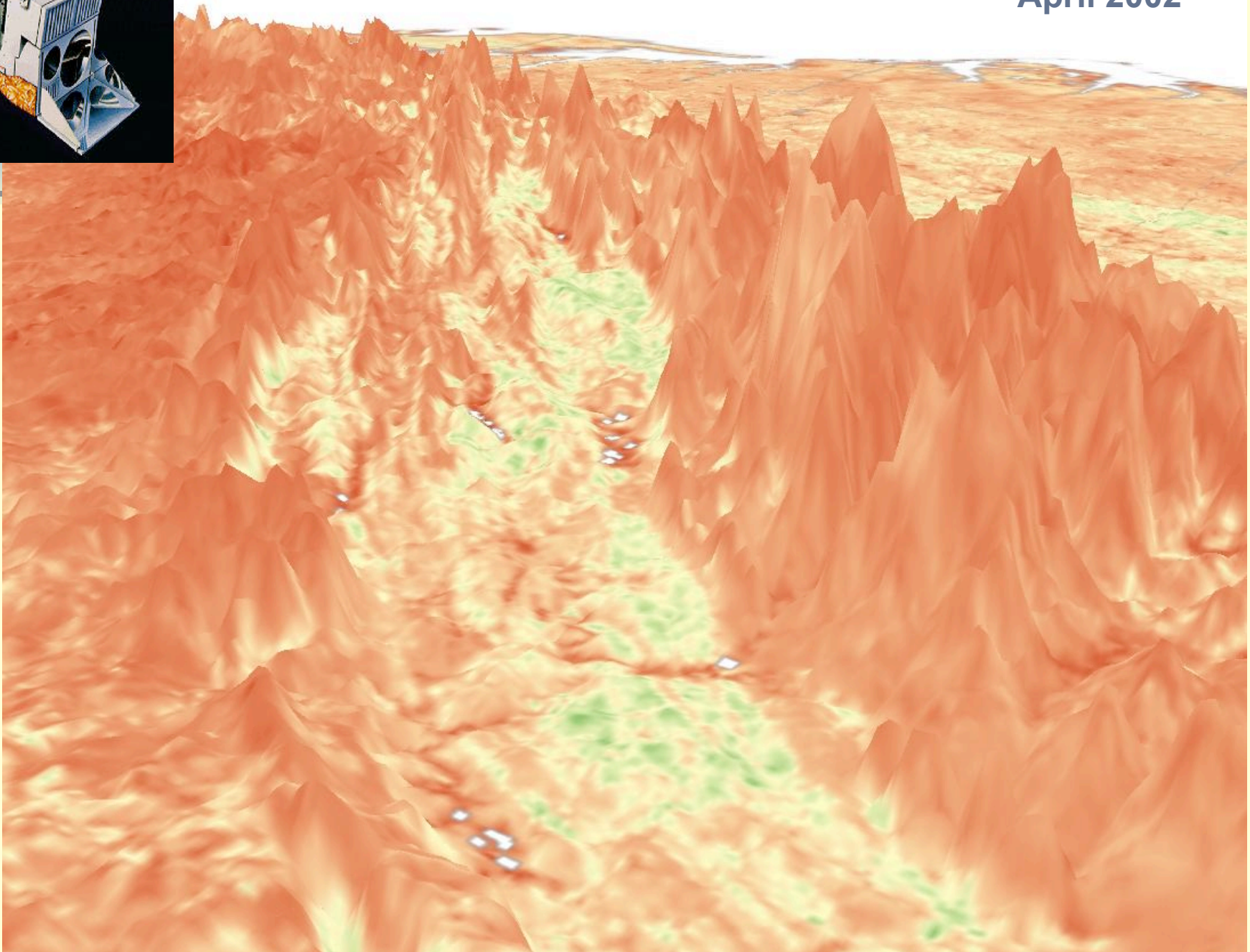
Bottom of King's Canyon

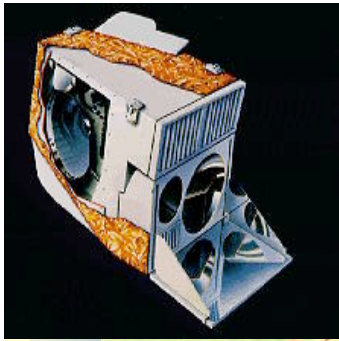




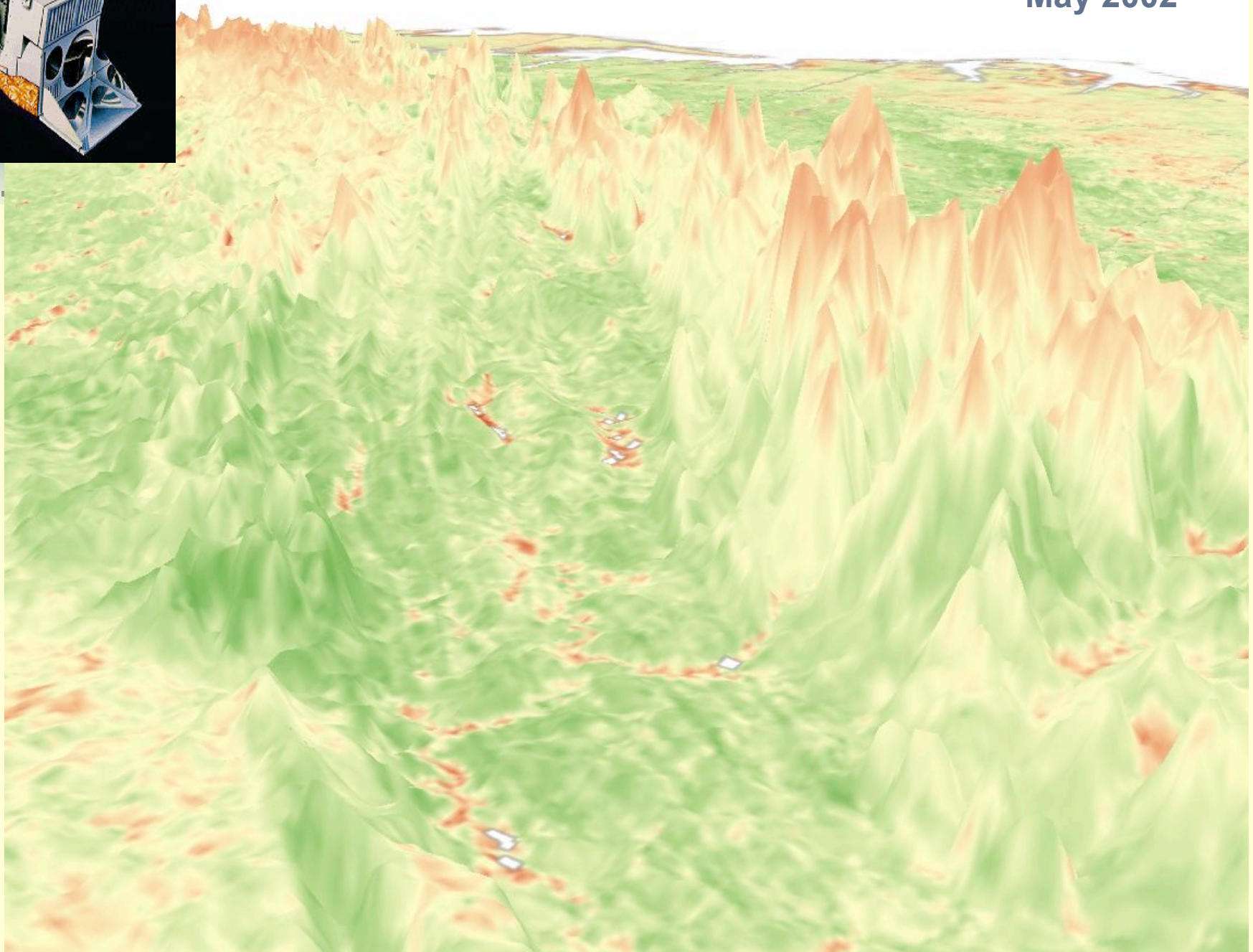


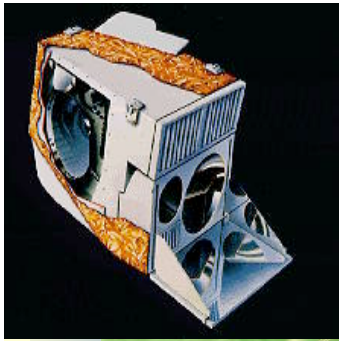
April 2002



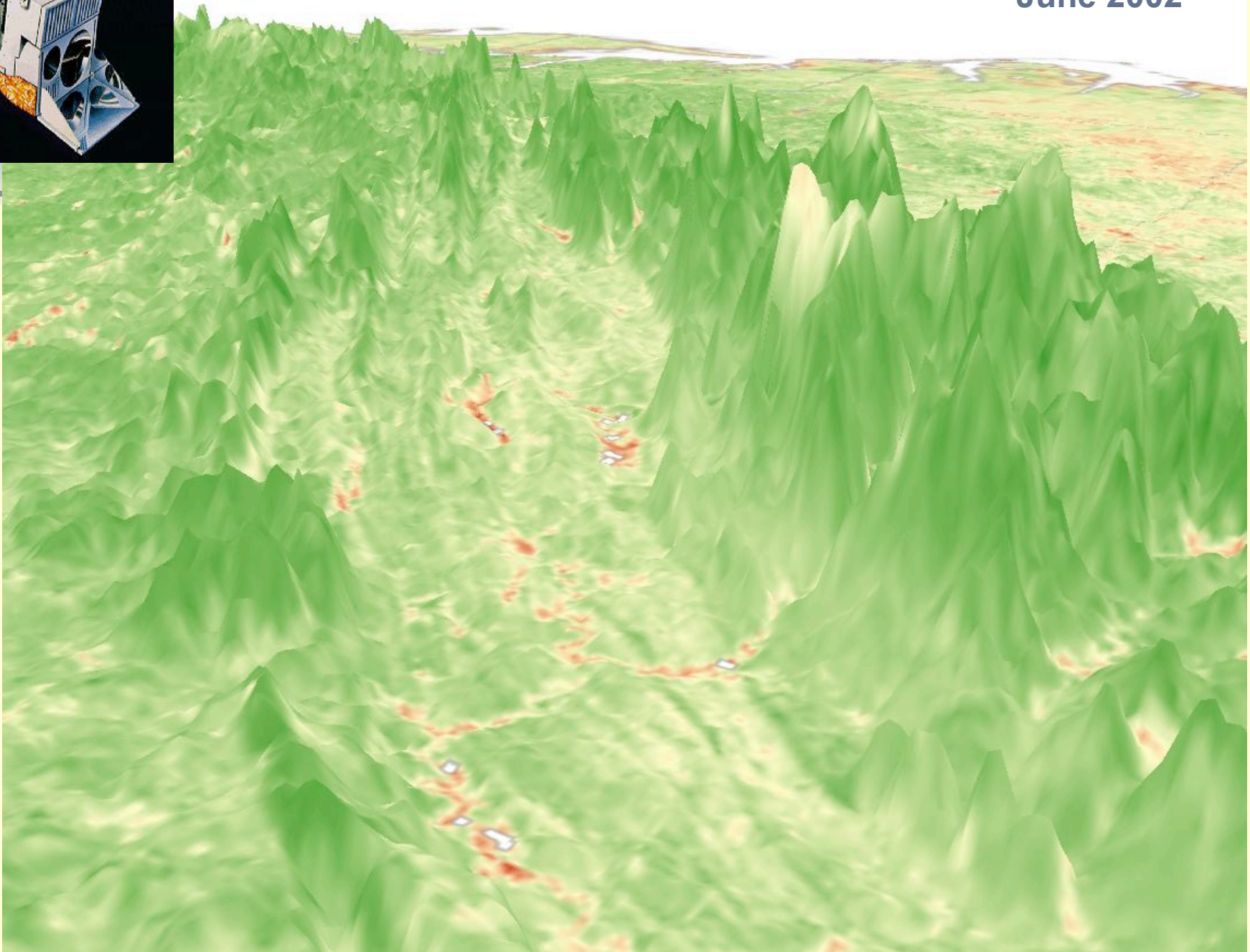


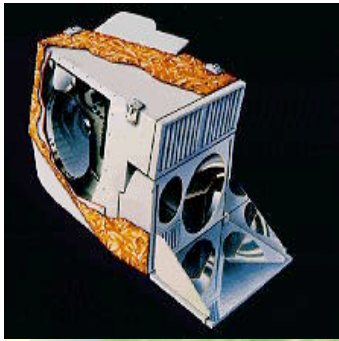
May 2002



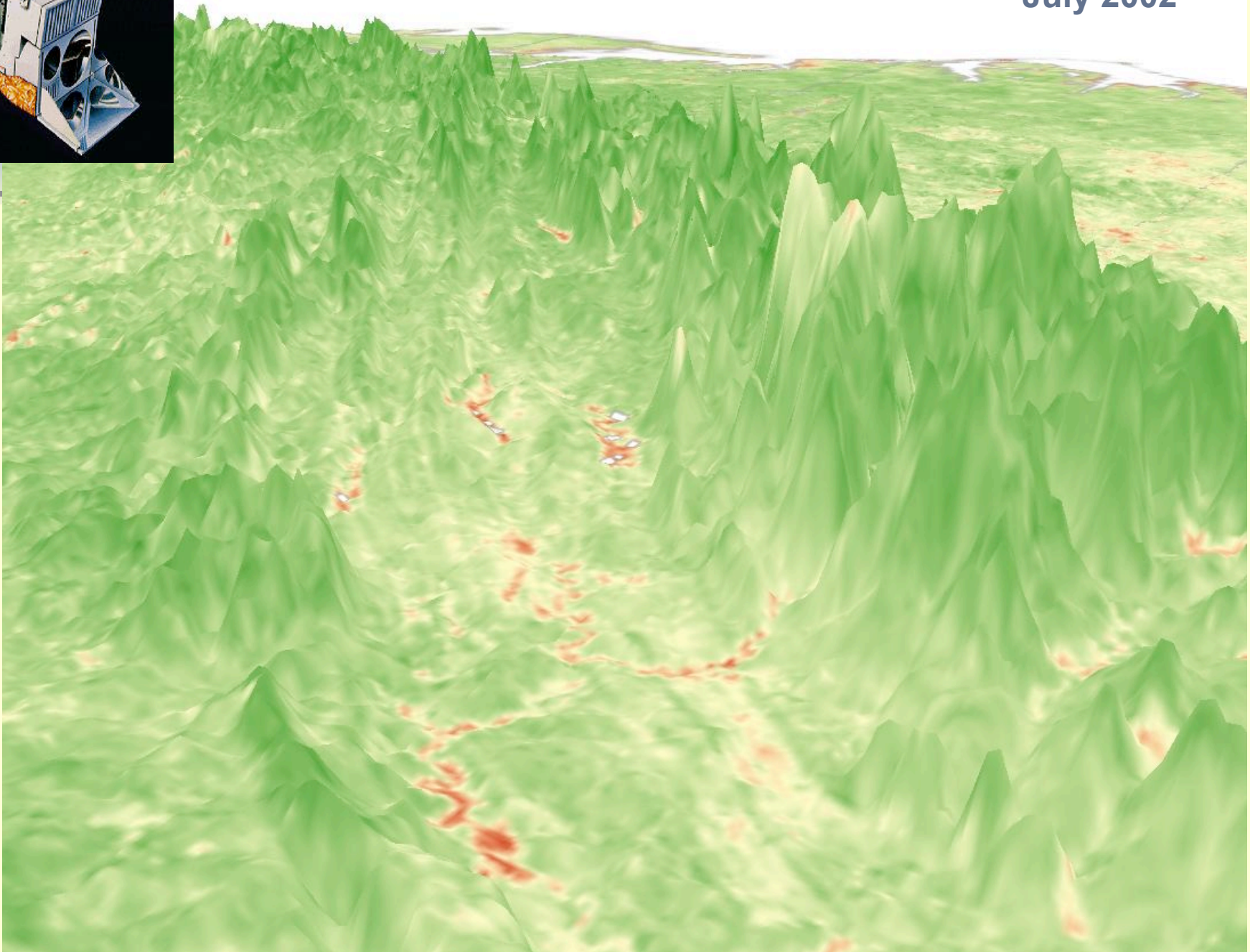


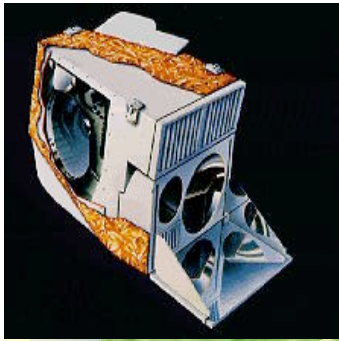
June 2002



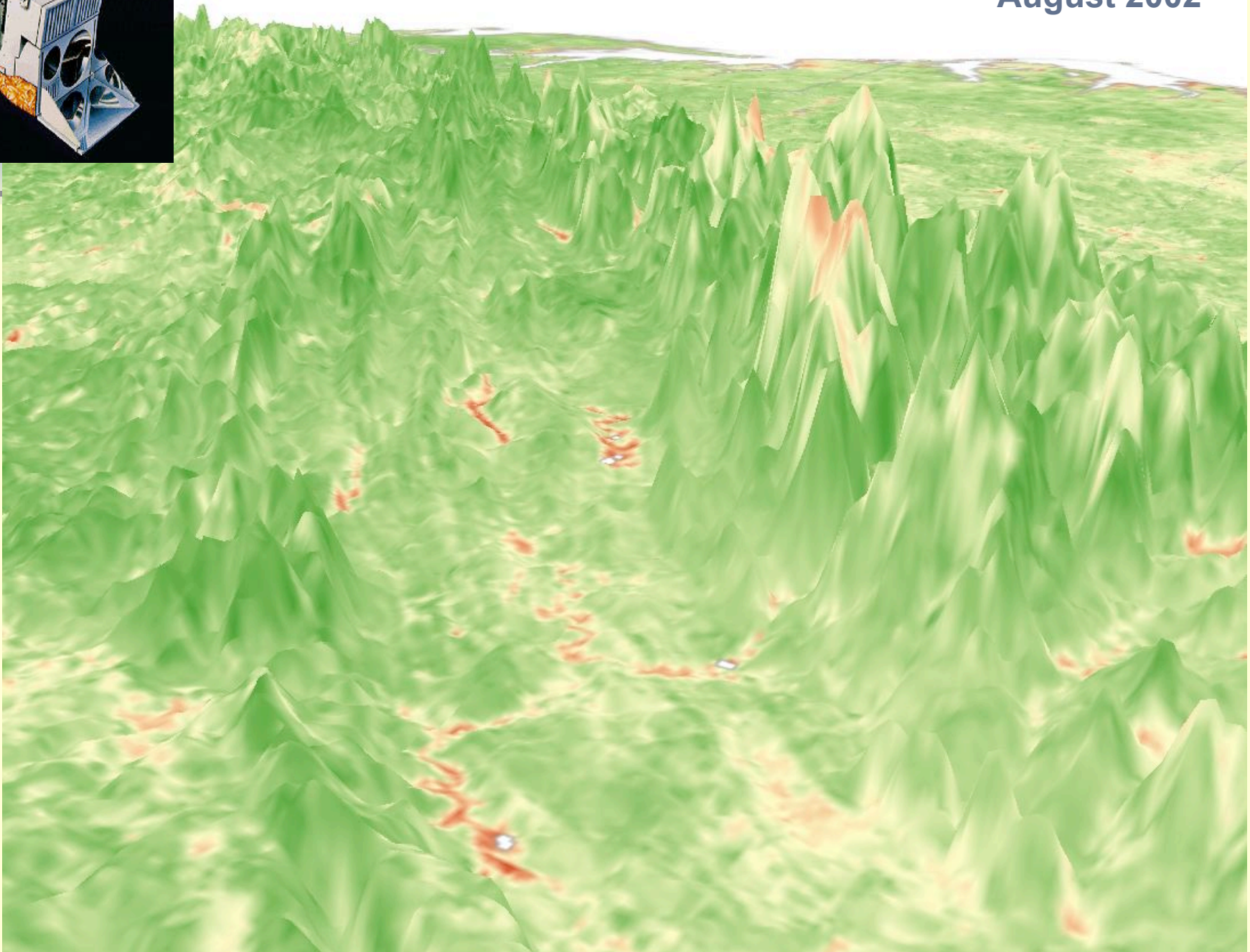


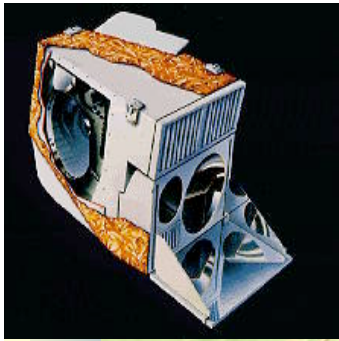
July 2002



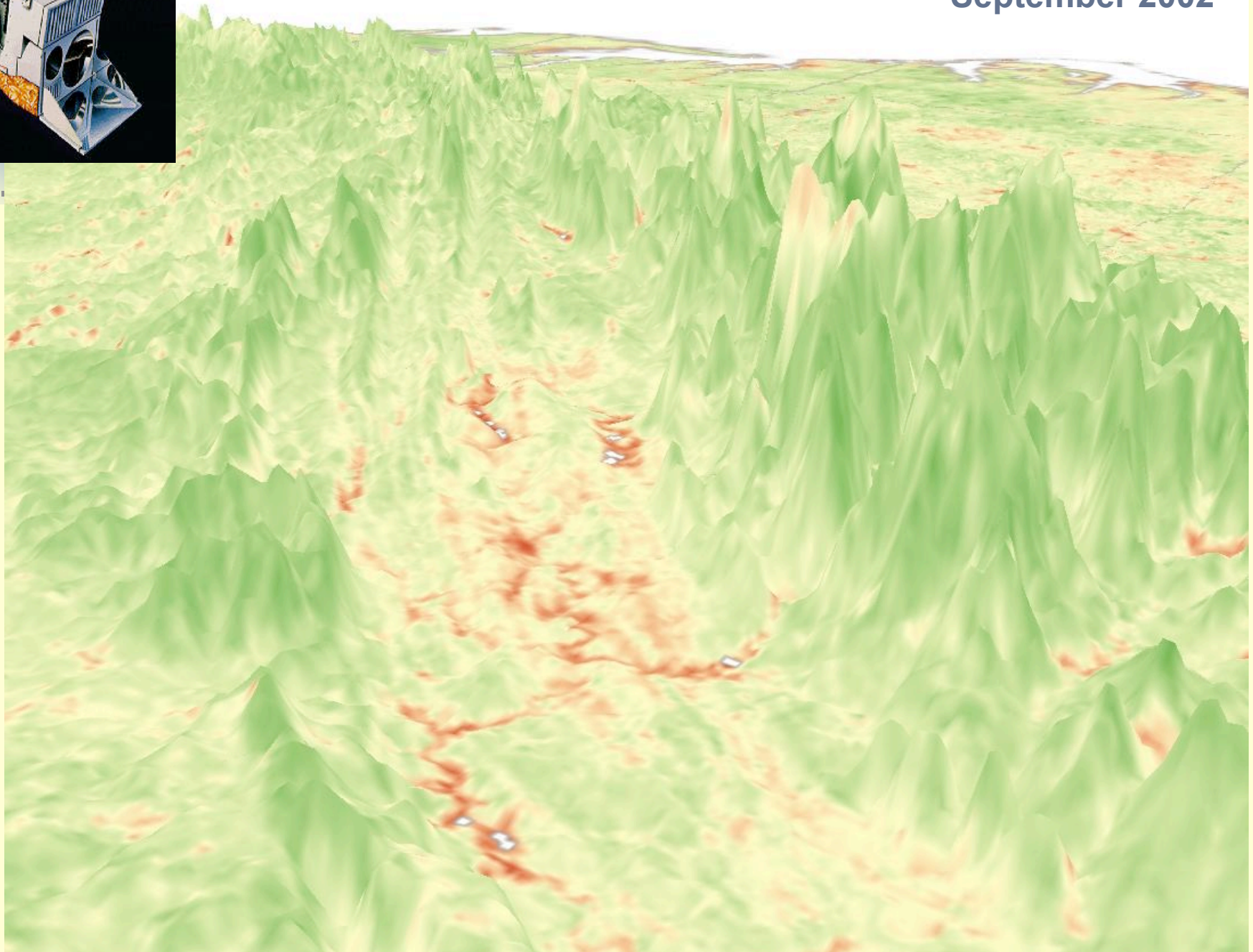


August 2002

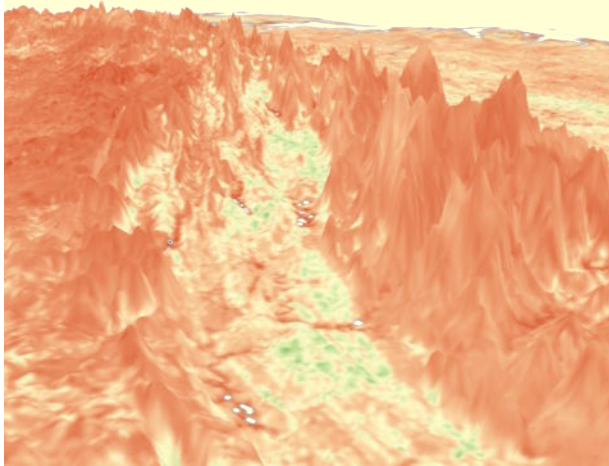




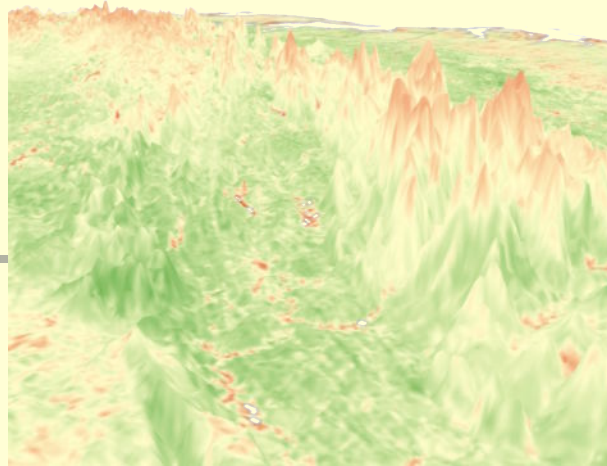
September 2002



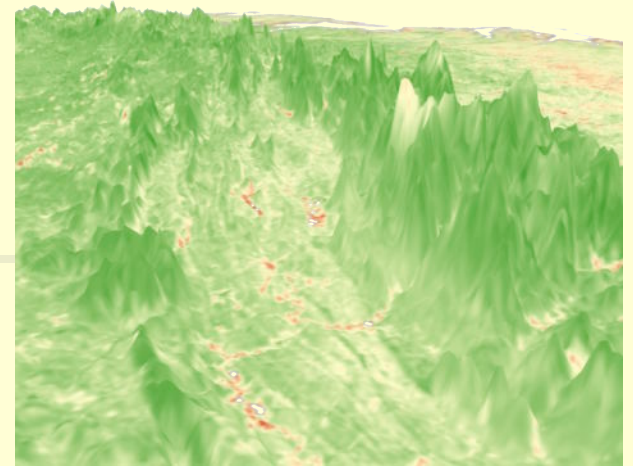
April 2002



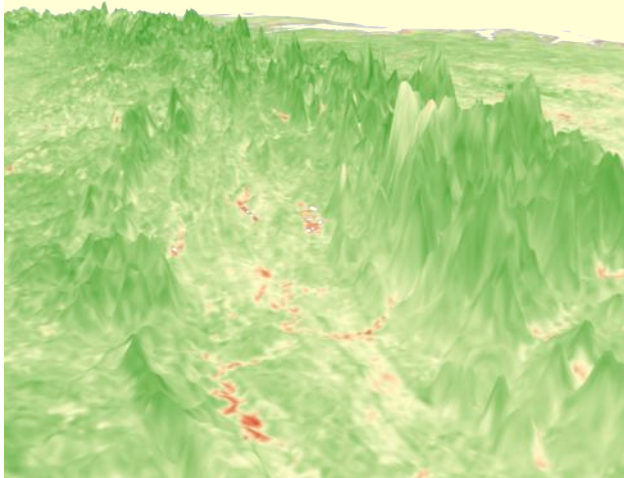
May 2002



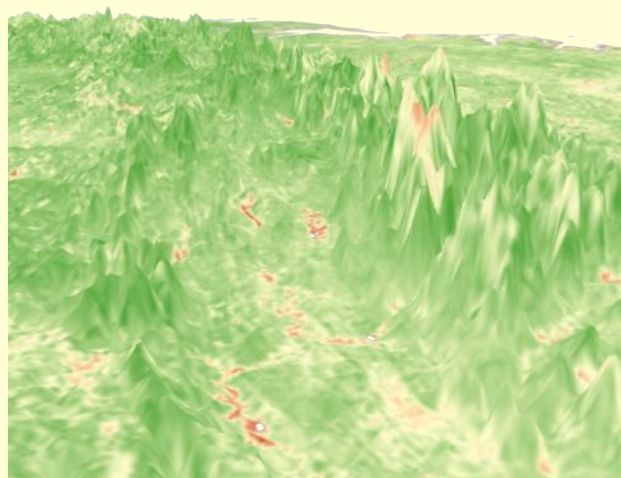
June 2002



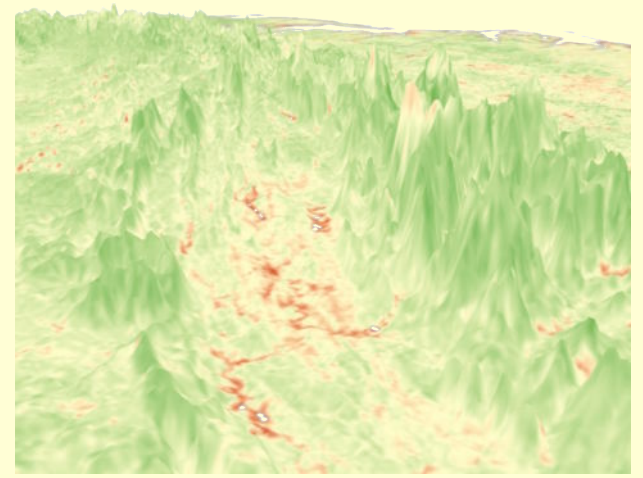
July 2002



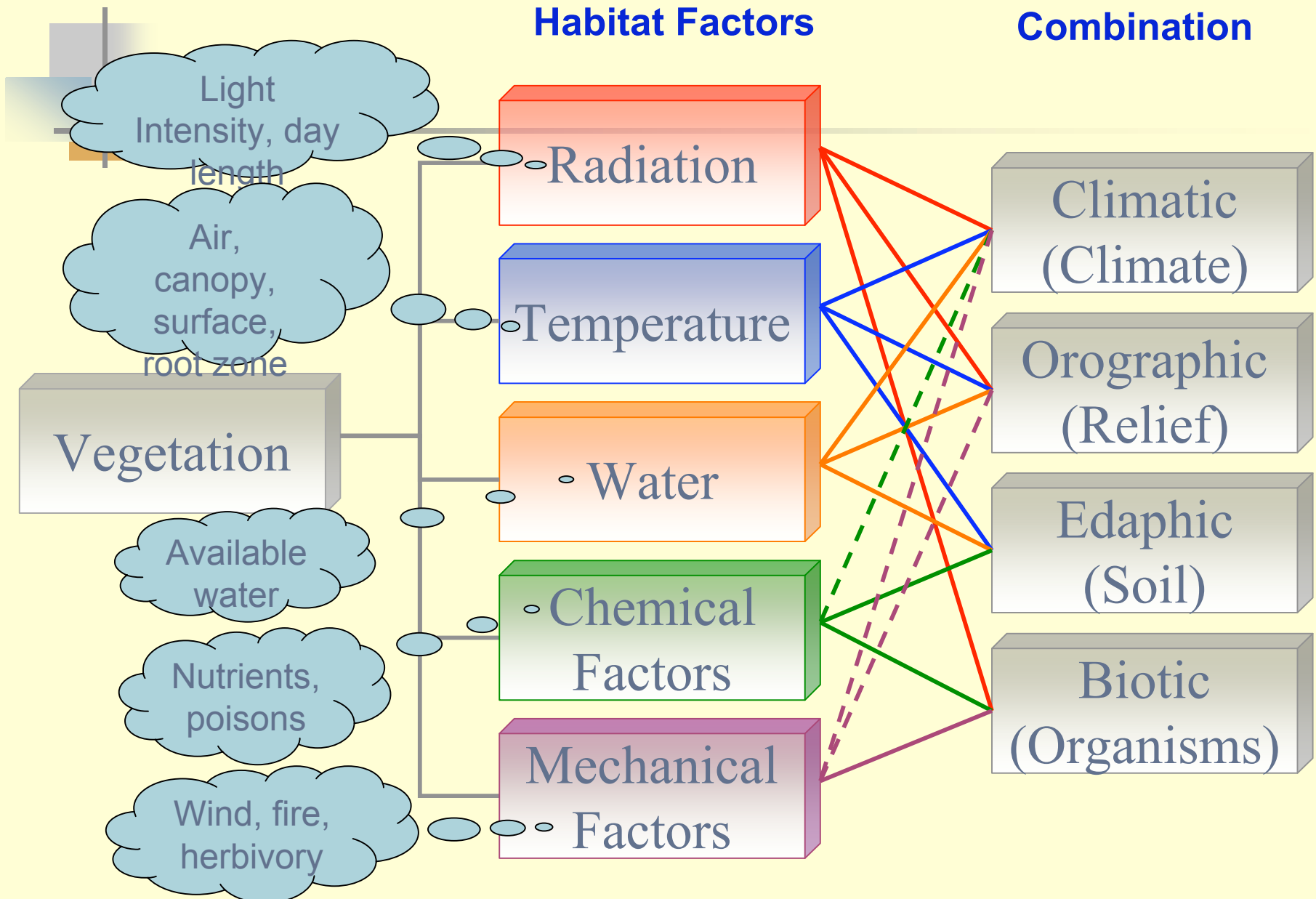
August 2002



September 2002



Complex Interactions



INPUTS

Elevation

Slope

Aspect

Compound Topographic Index

Distance to Nearest Stream

Percent Sand

Percent Silt

Percent Clay

Land Use/Land Cover

Possibly many more ...



OUTPUT

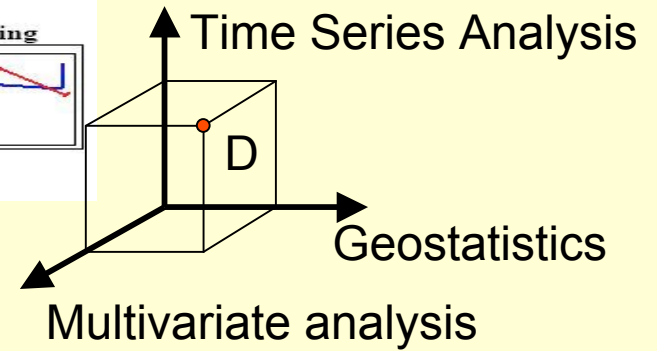
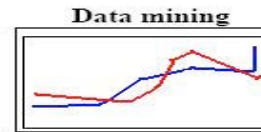
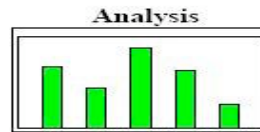
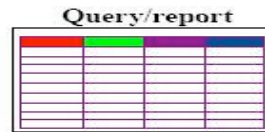
April 2000

April 2001

April 2002

April 2003

Average
Enhanced
Vegetation
Index for April

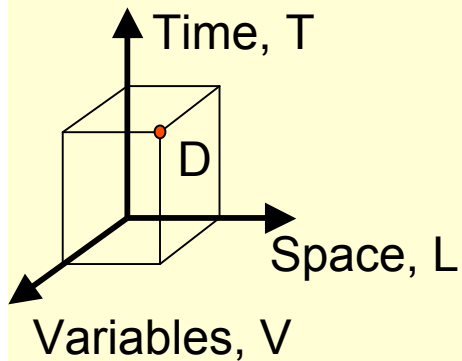
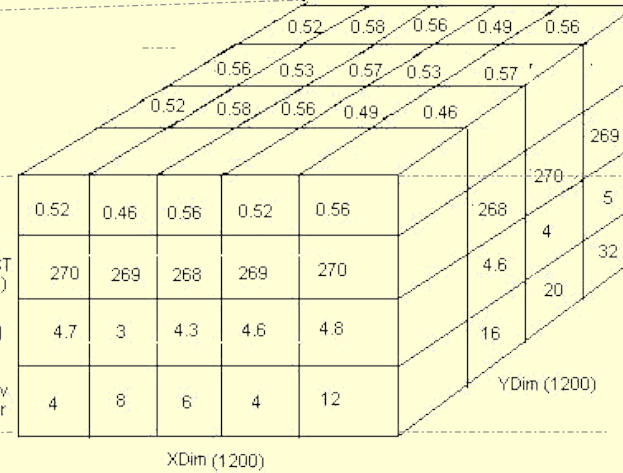
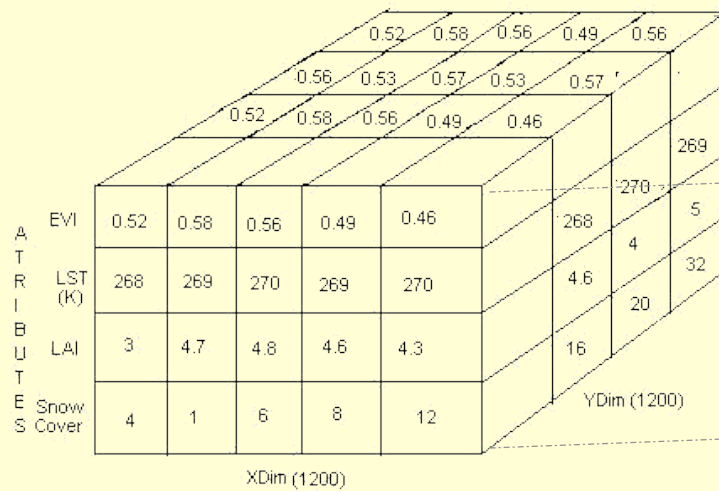


Data to Knowledge

D2K

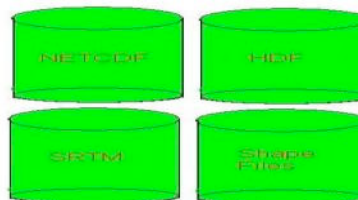
Jan

Feb

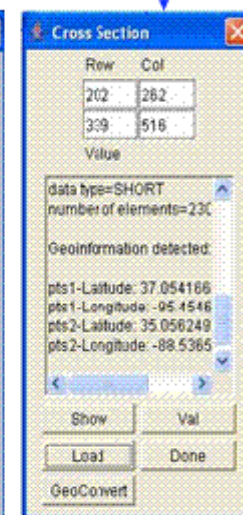
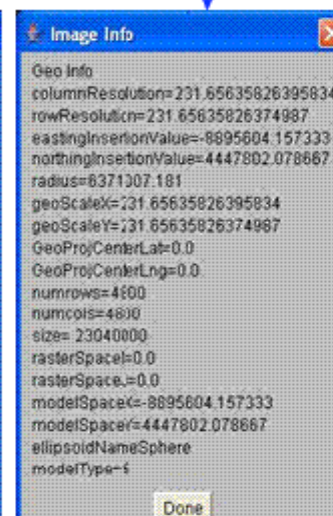
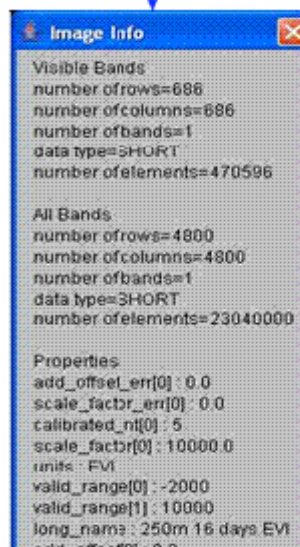
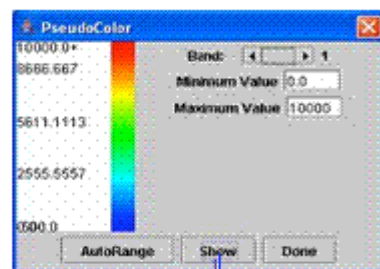
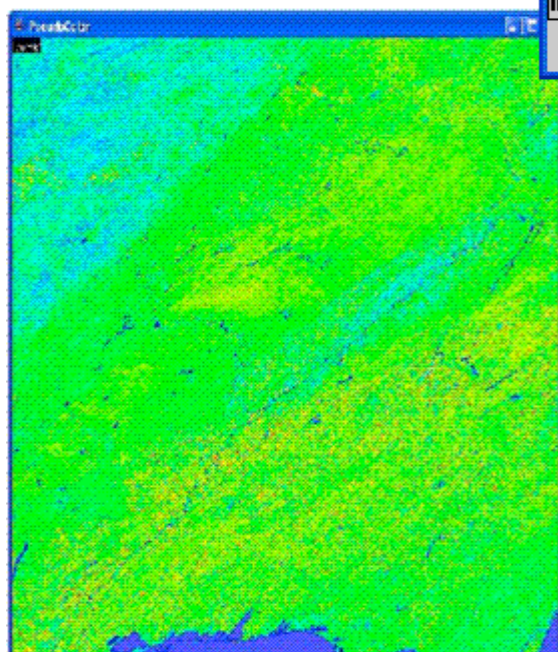
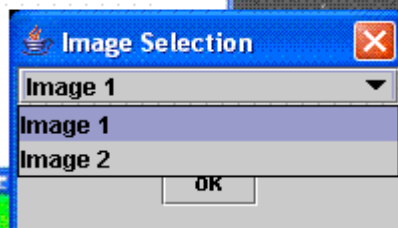
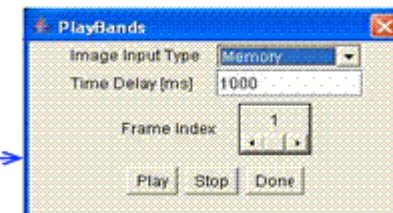
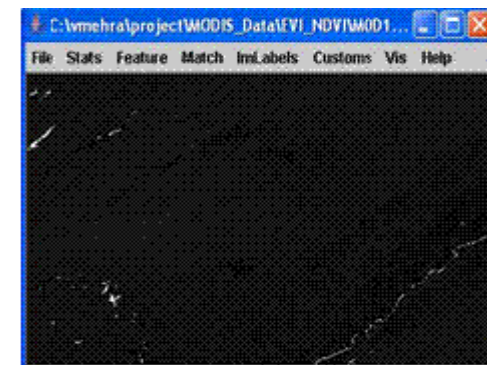
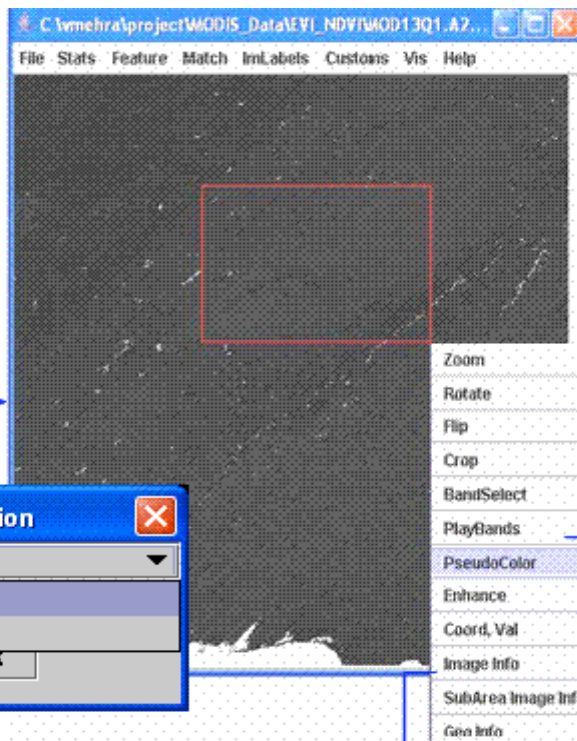
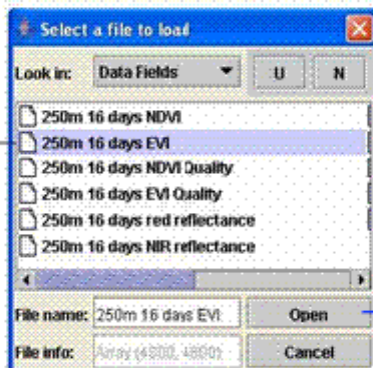
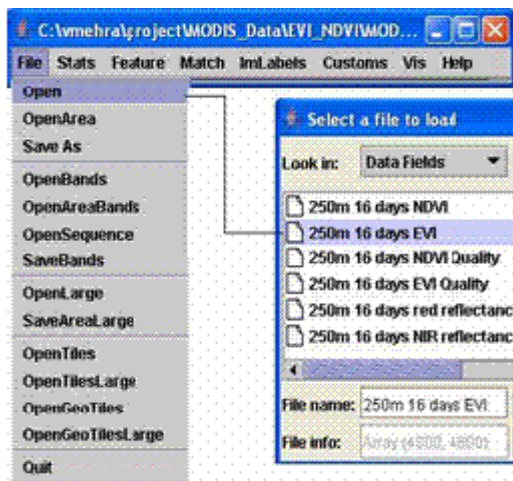


i2K

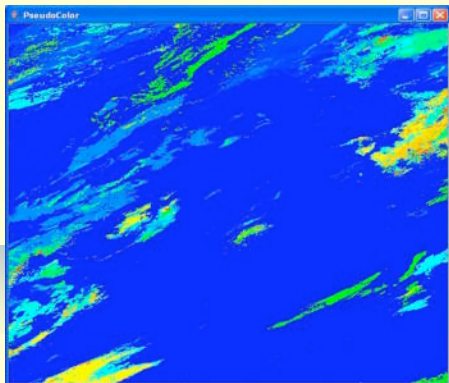
Image to Knowledge



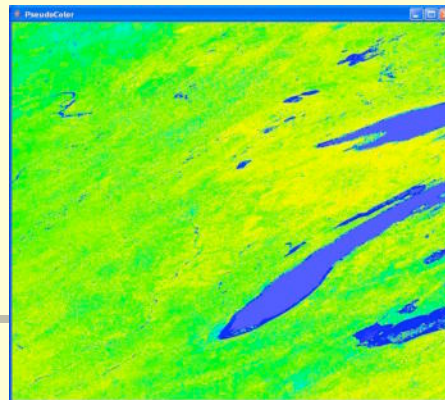
Data Files



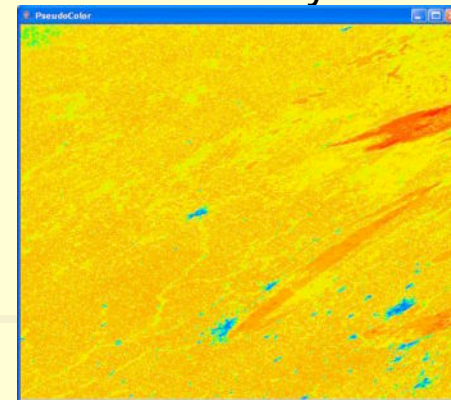
Snow Cover



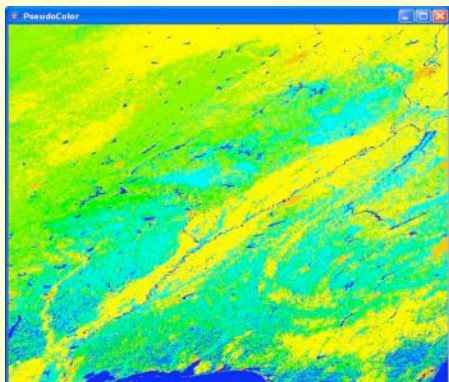
NDVI



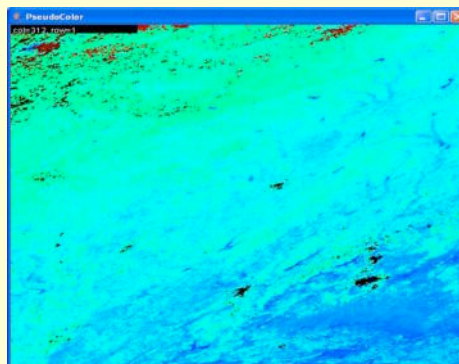
Emissivity



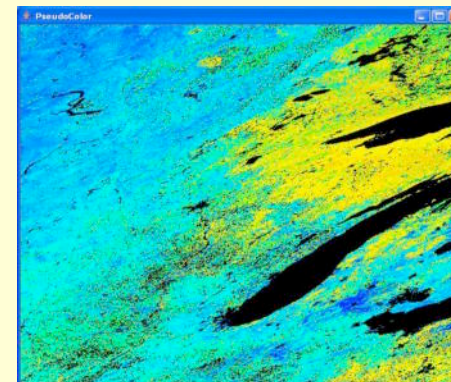
Land Cover



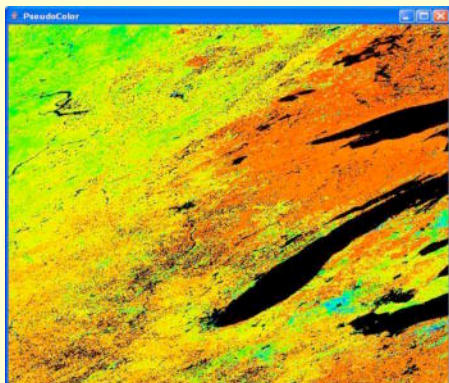
Albedo ($\lambda = 648\text{nm}$)



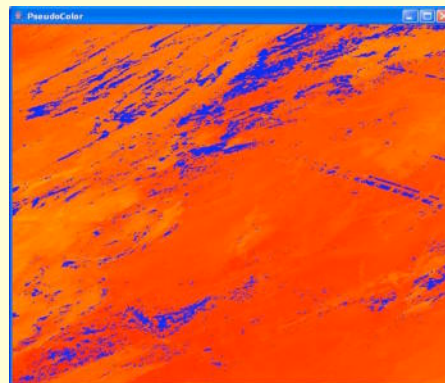
LAI



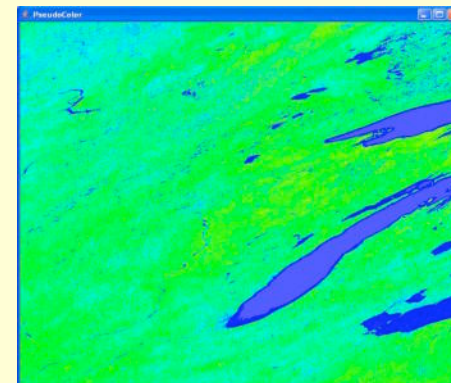
FPAR



LST

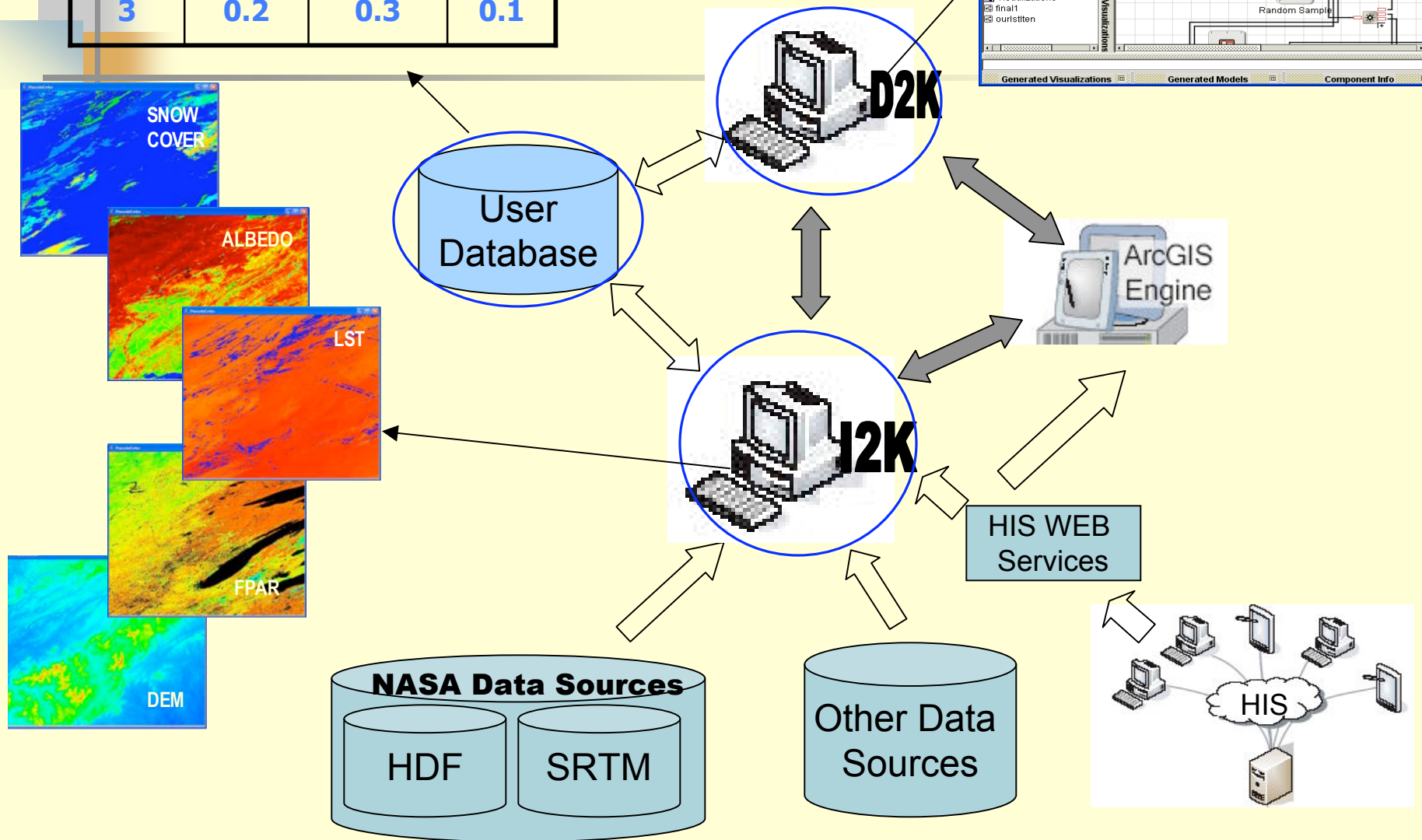
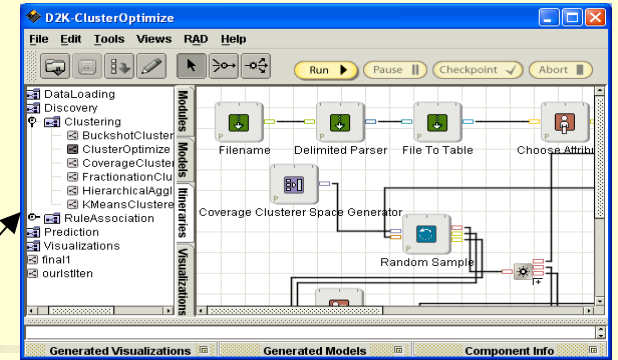


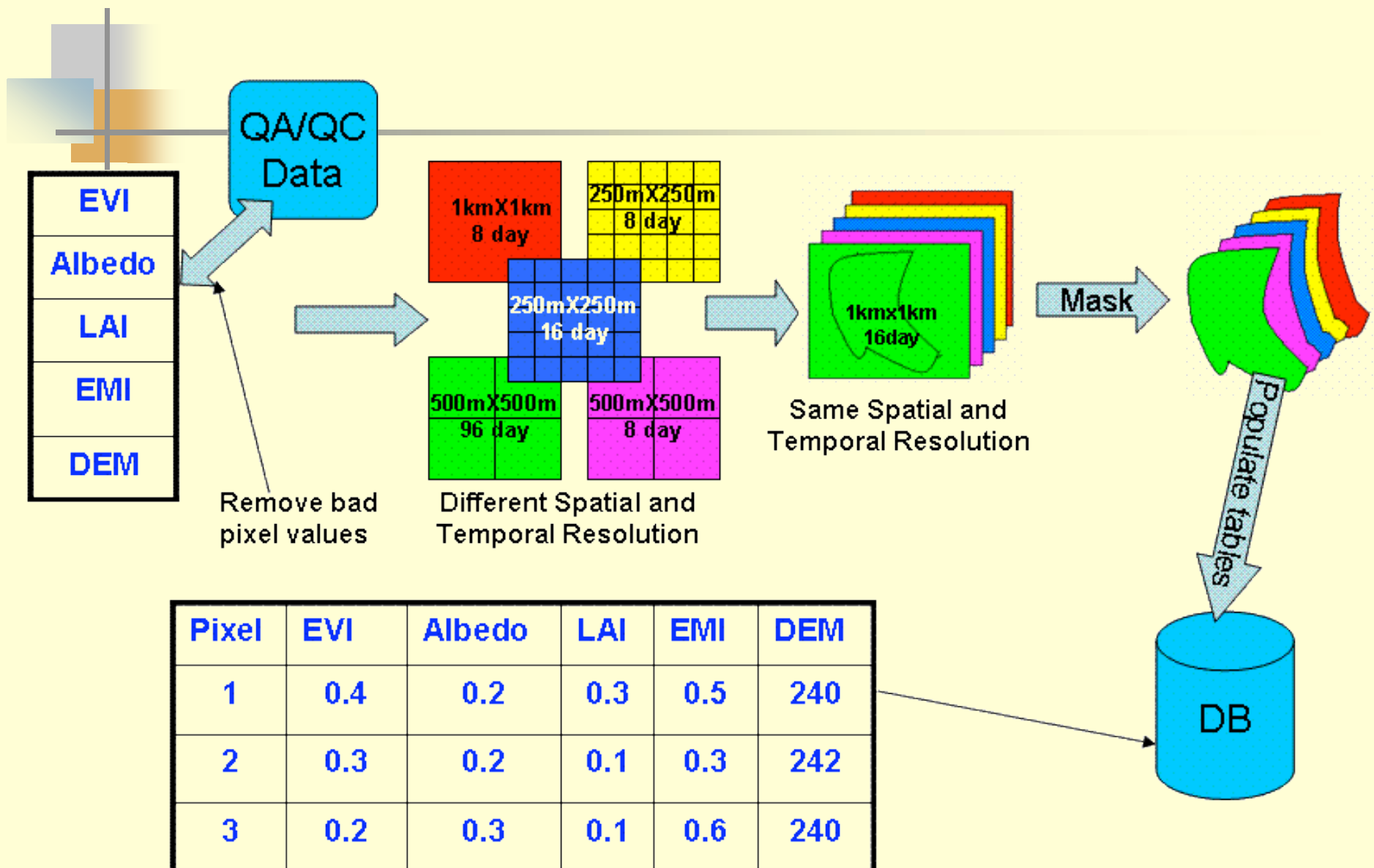
EVI

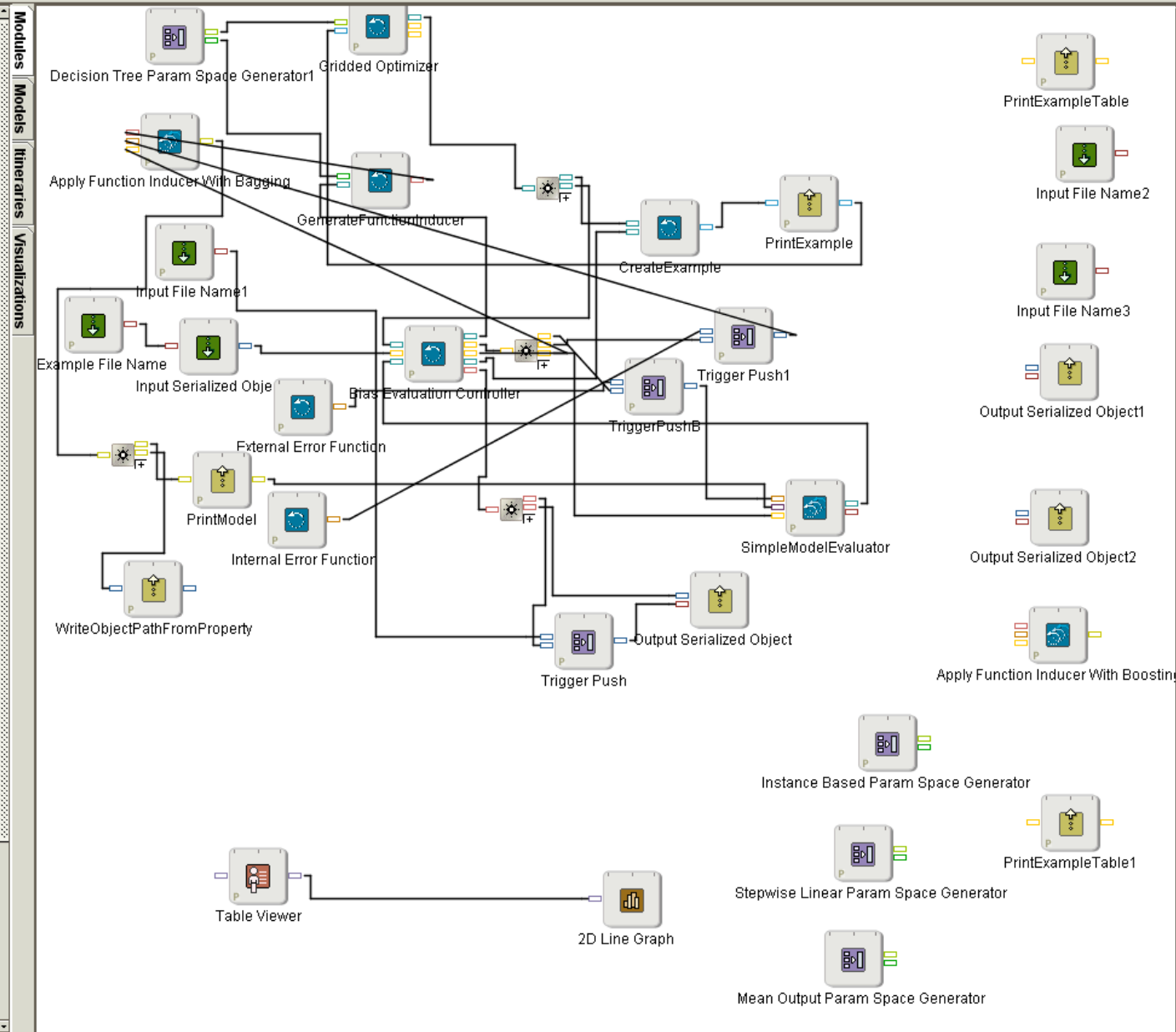
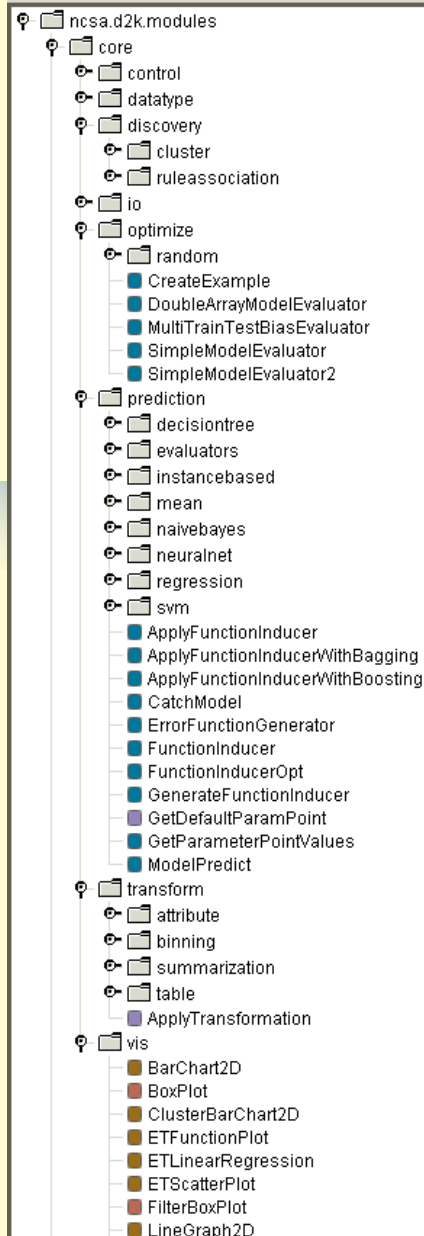


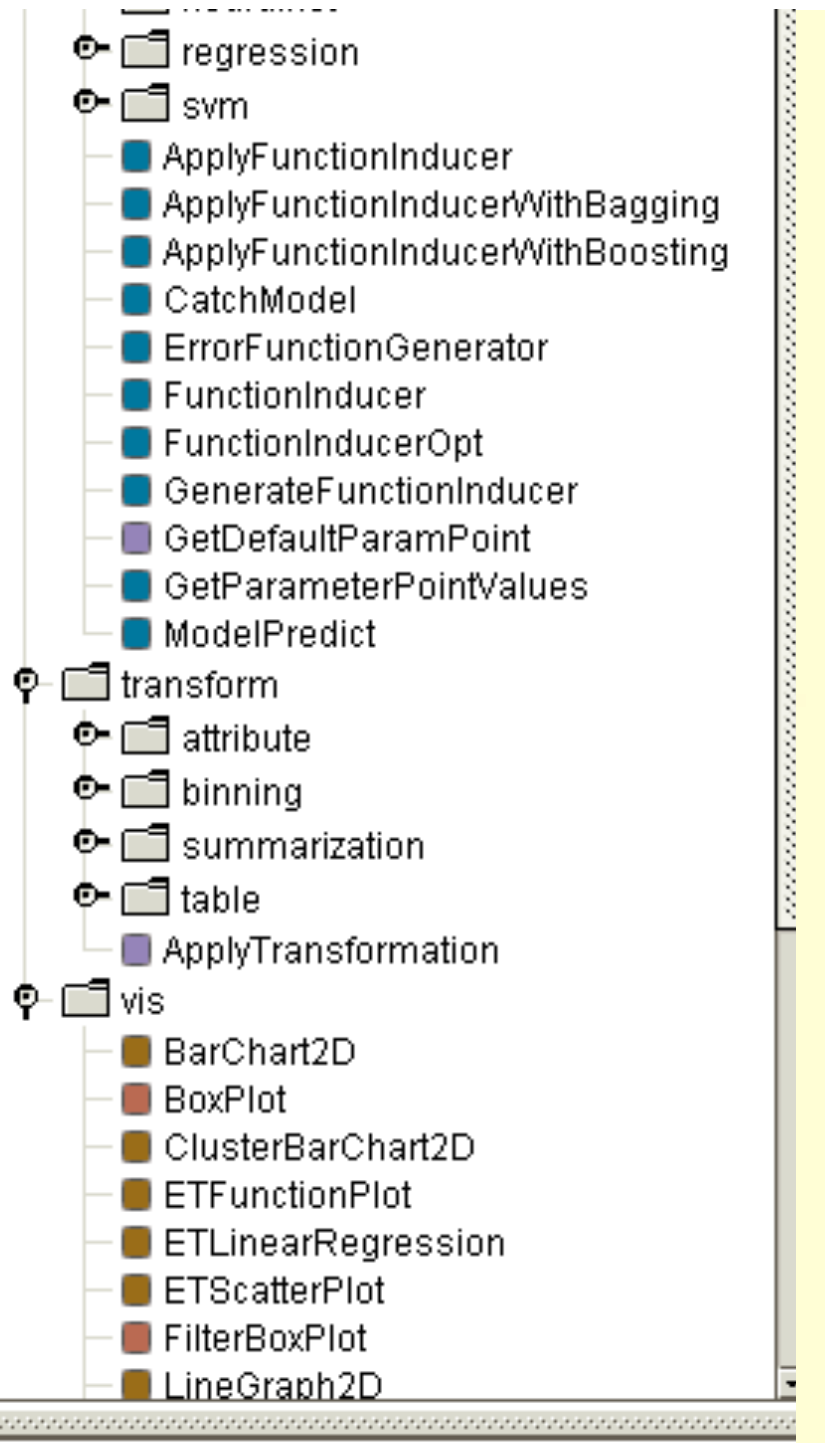
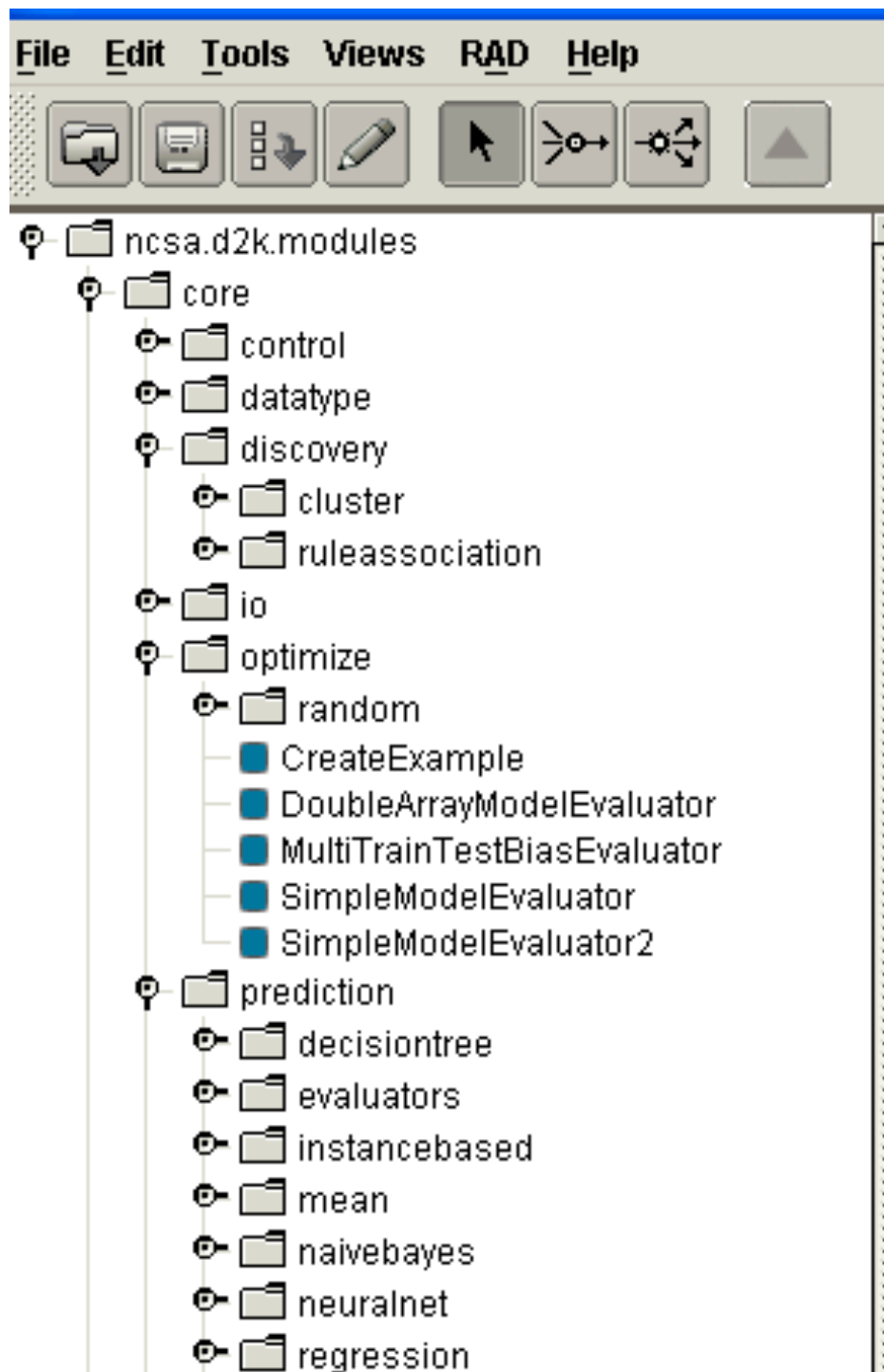
Pixel	EVI	Albedo	LAI
1	0.4	0.2	0.3
2	0.3	0.2	0.1
3	0.2	0.3	0.1

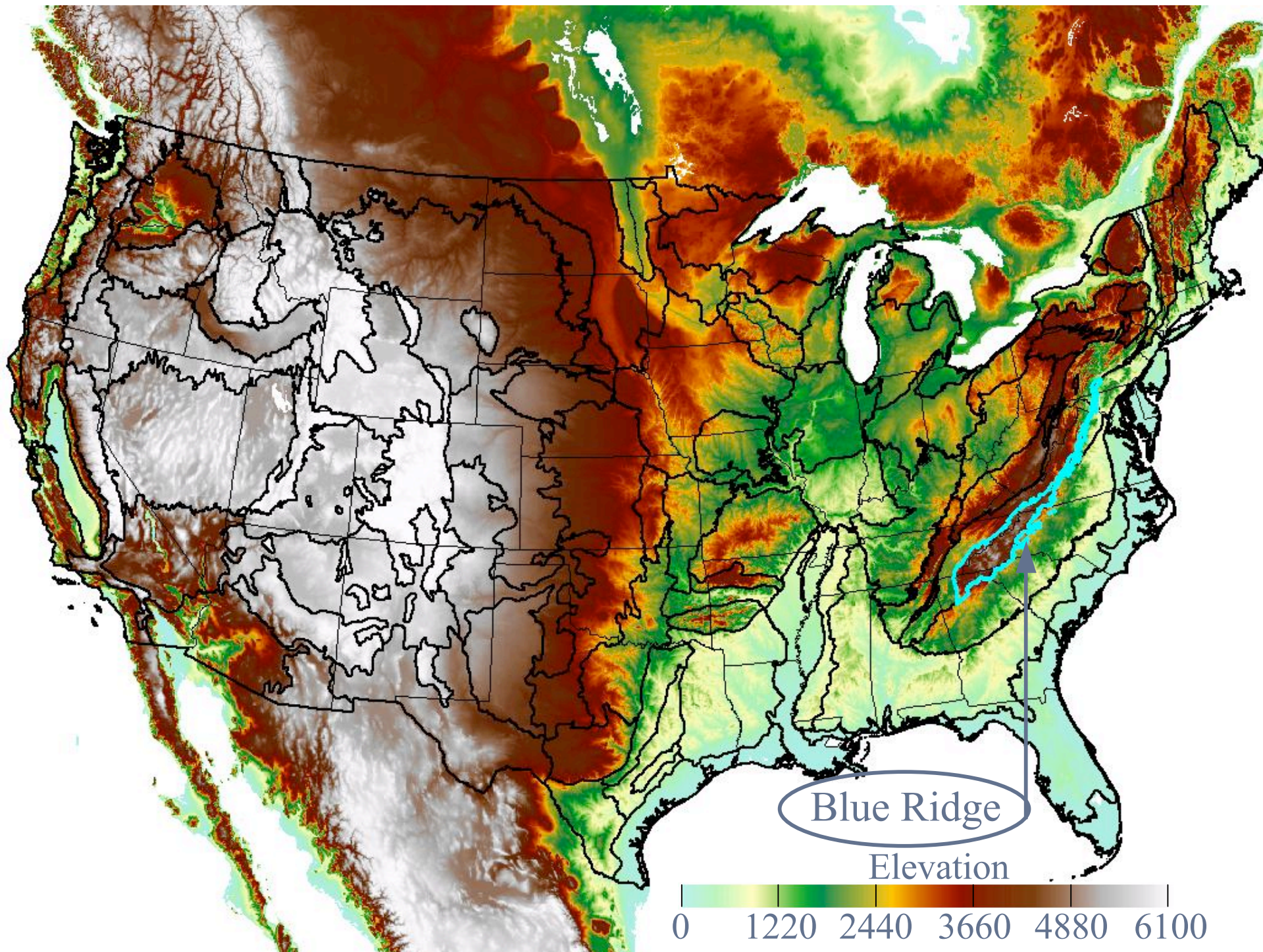
System Architecture

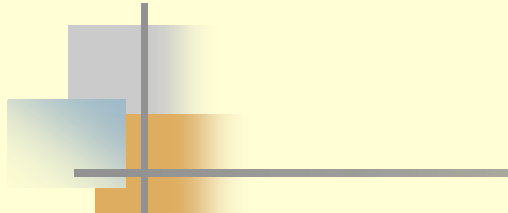















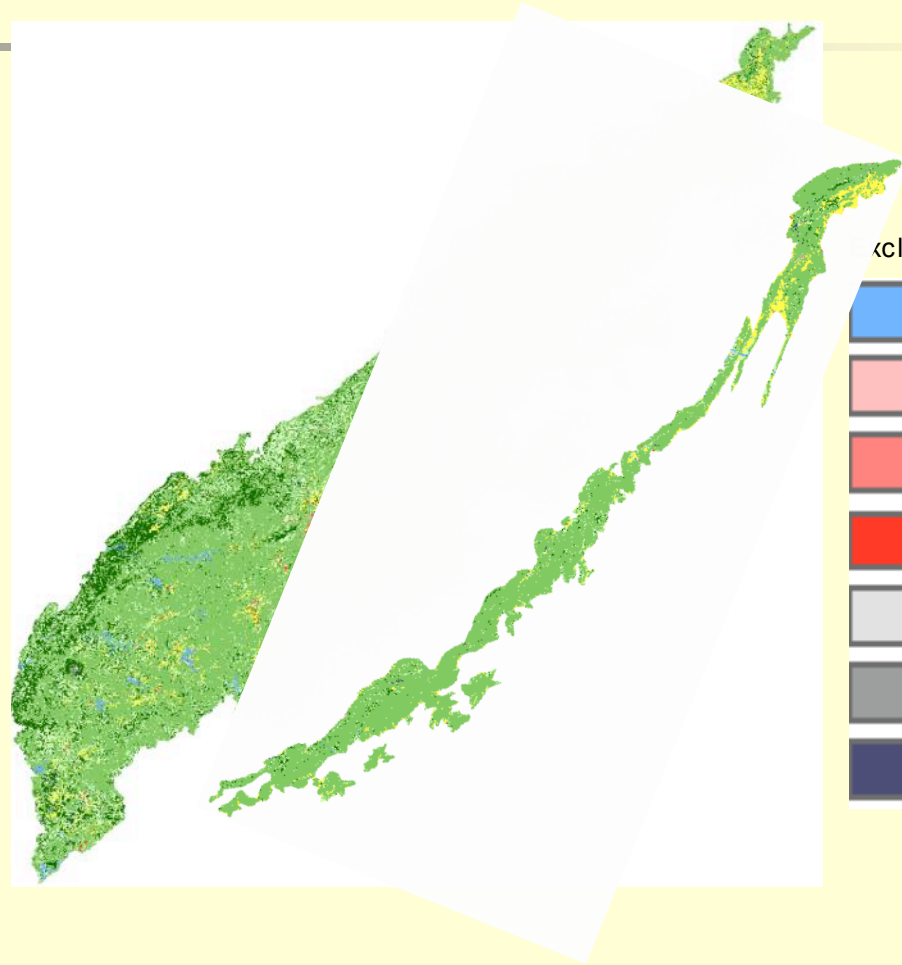






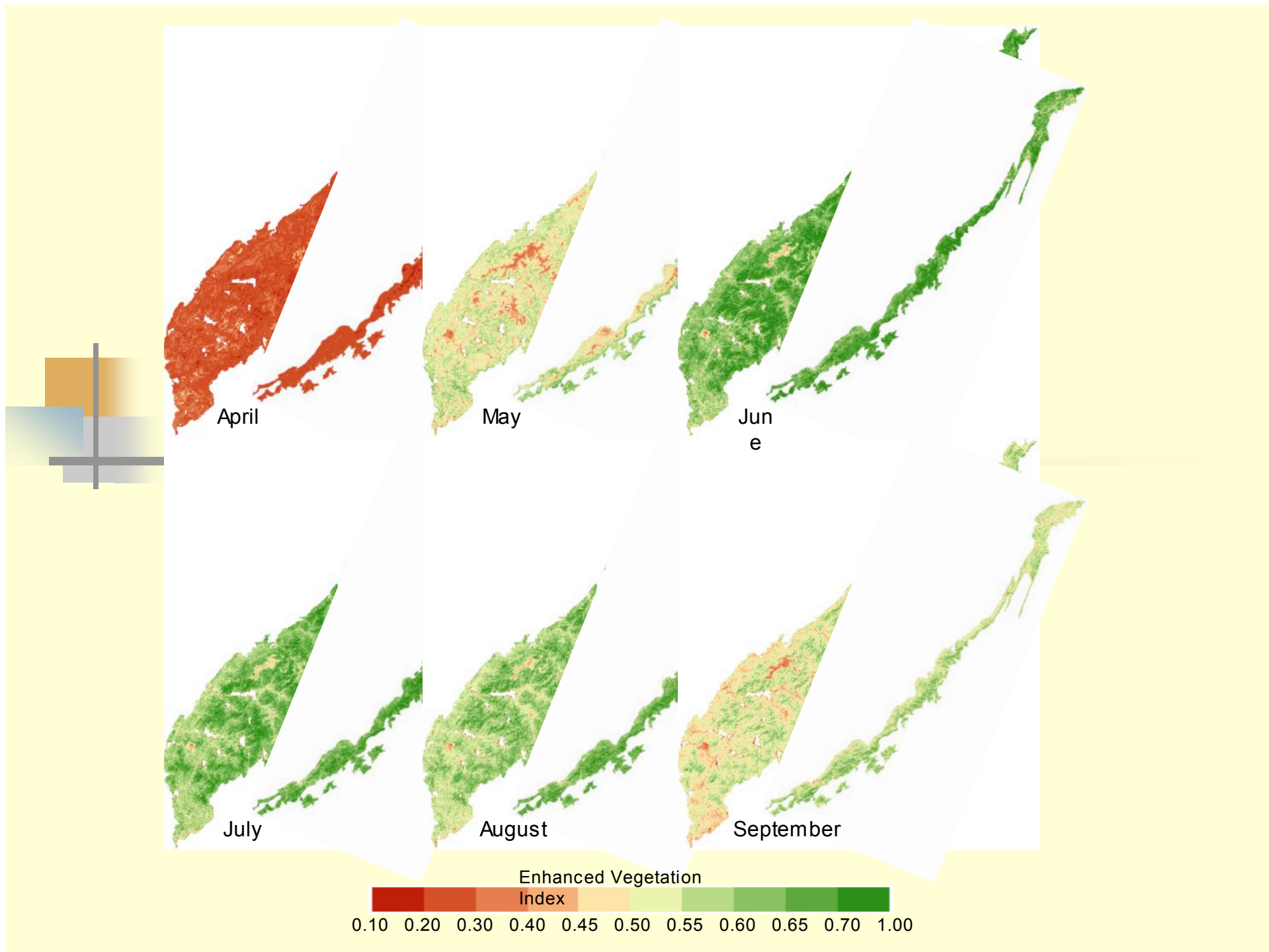
Included Land Covers

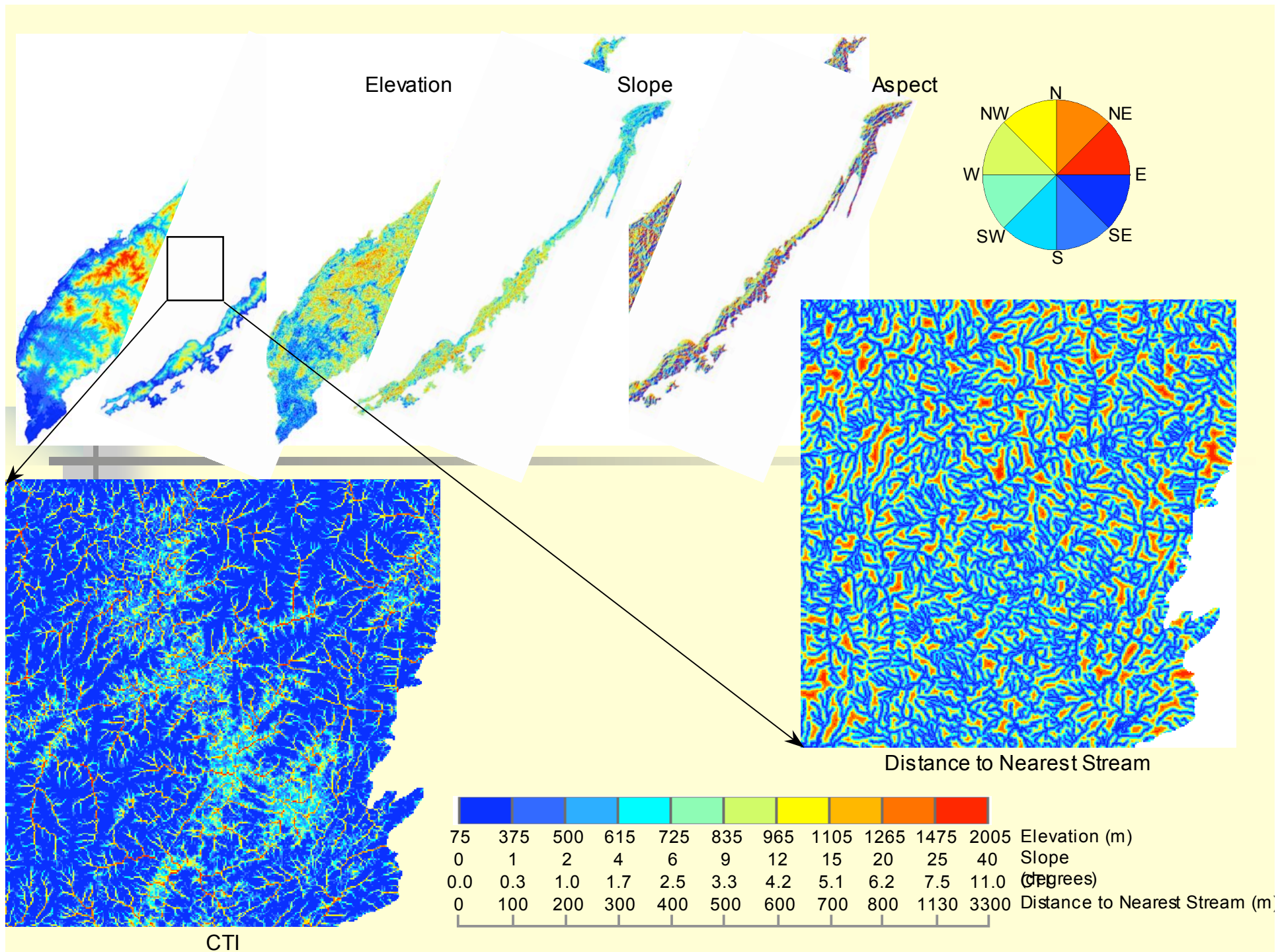
-  Deciduous Forest
-  Evergreen Forest
-  Mixed Forest
-  Pasture/Hay
-  Row Crops
-  Small Grains
-  Urban Grasses
-  Woody Wetlands
-  Herbaceous Wetlands

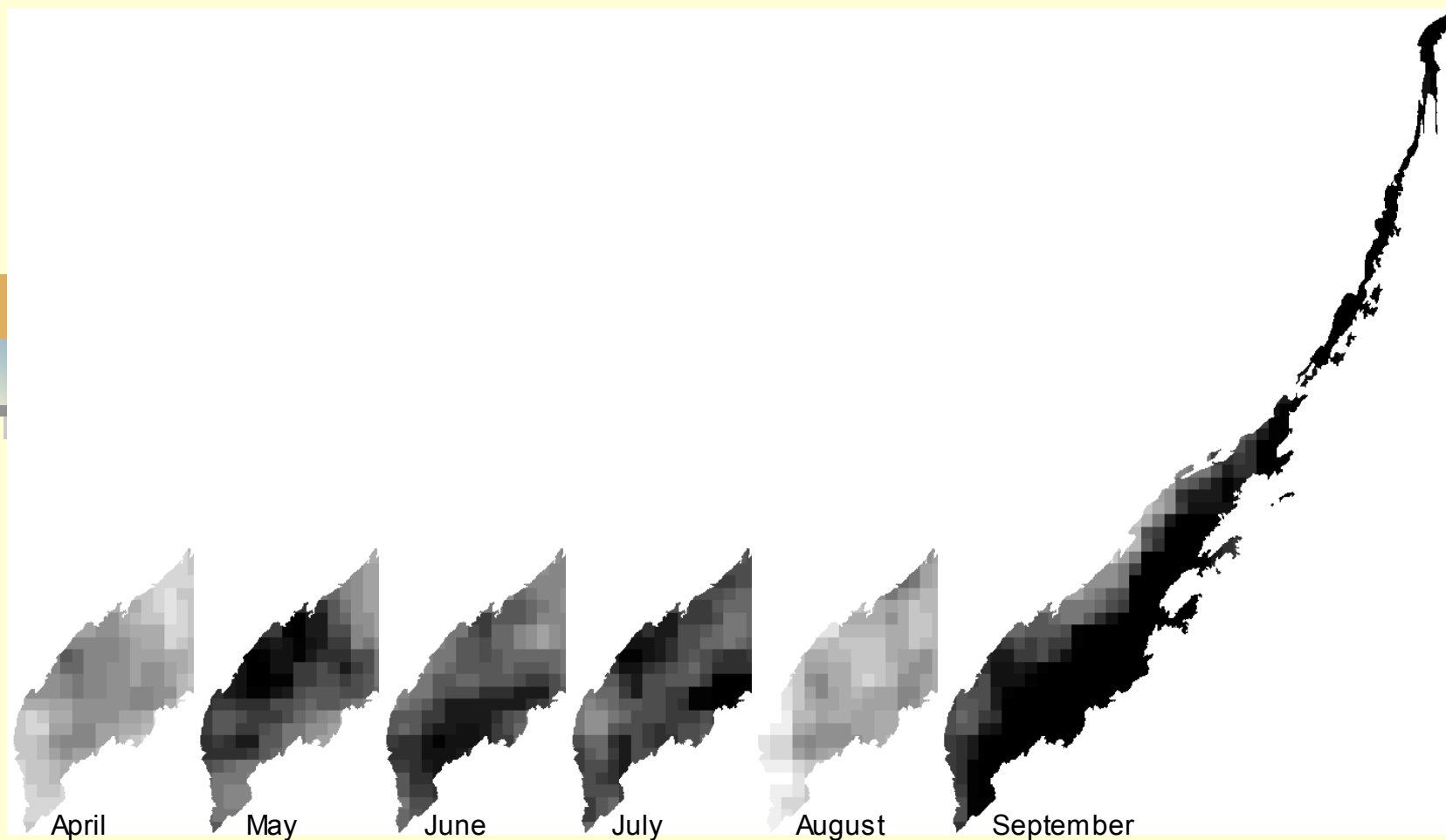


Excluded Land Covers

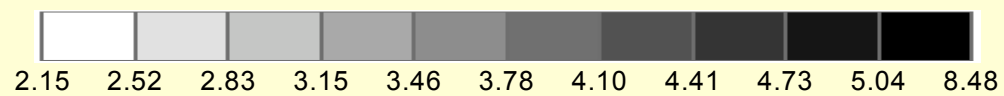
-  Water
-  Low Intensity Residential
-  High Intensity Residential
-  Commercial/Industry/Transportation
-  Bare Rock/Sand/Clay
-  Quarries/Strip Mines/Gravel Pits
-  Transitional

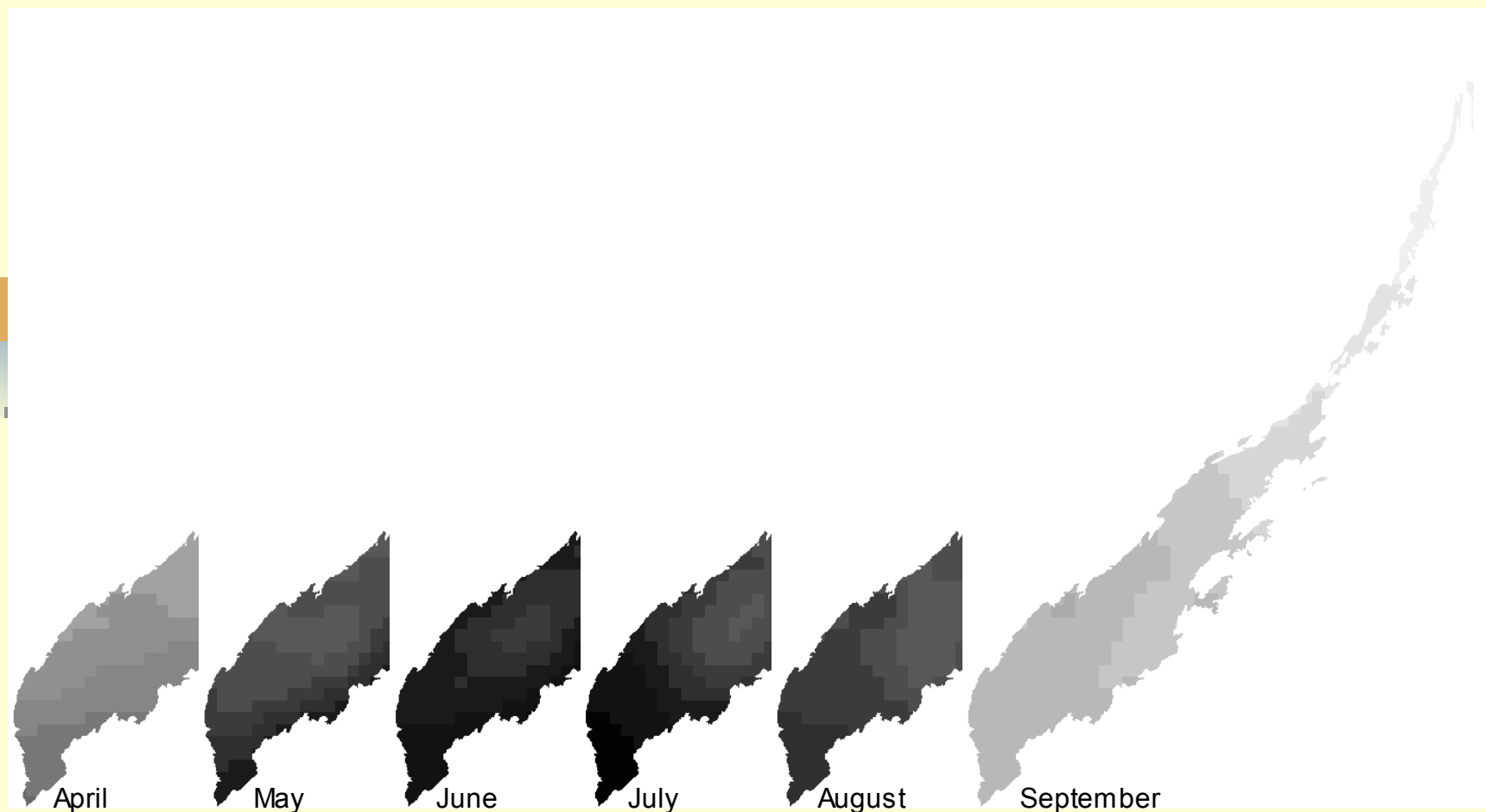




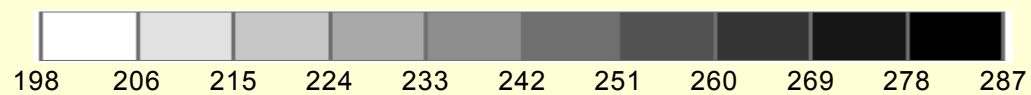


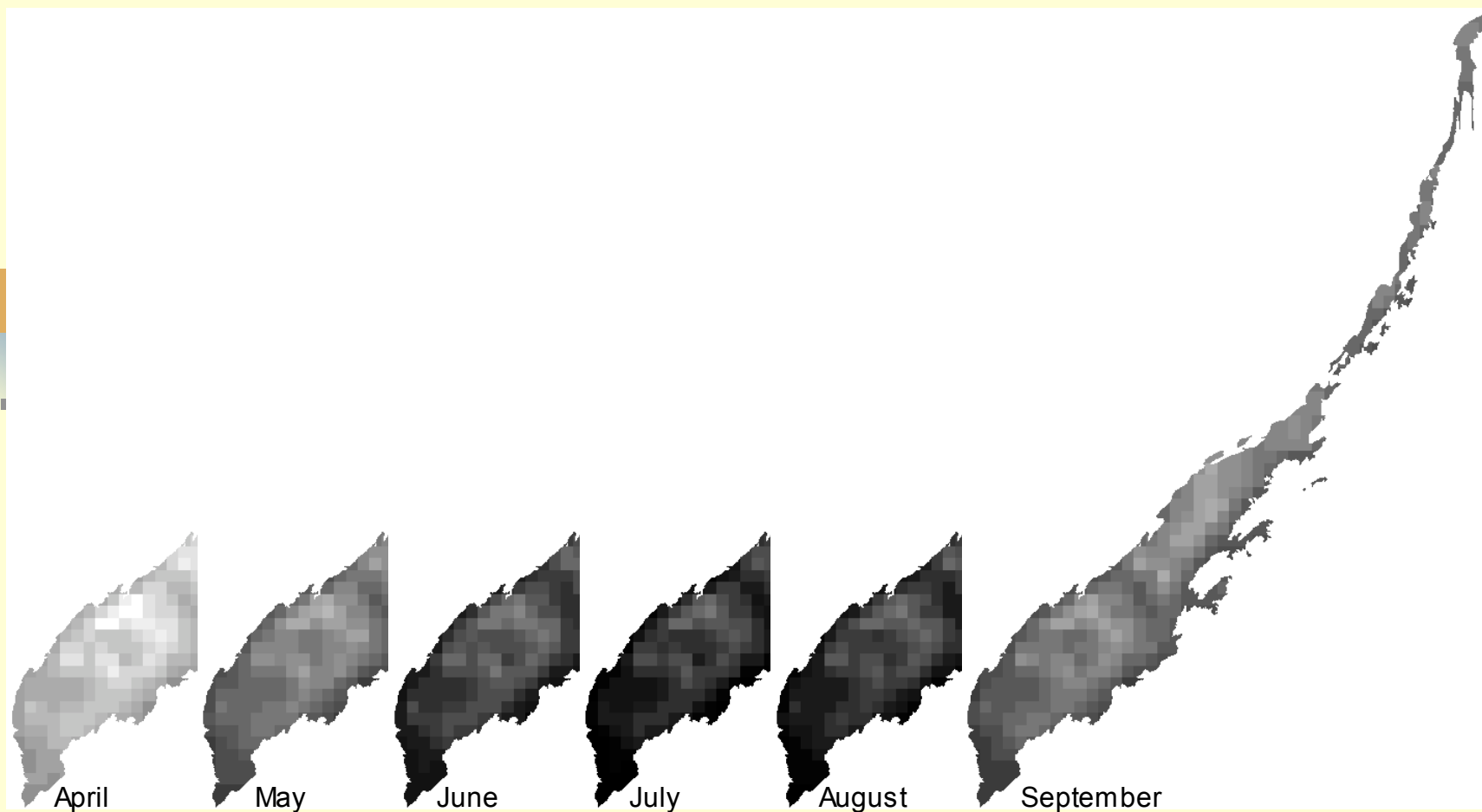
Precipitation (mm)



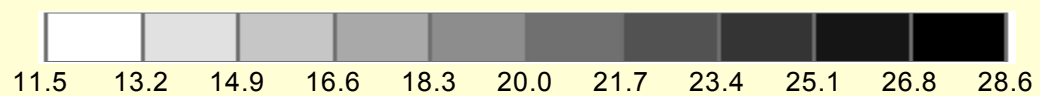


Incoming Shortwave Radiation Flux (W/m^2)



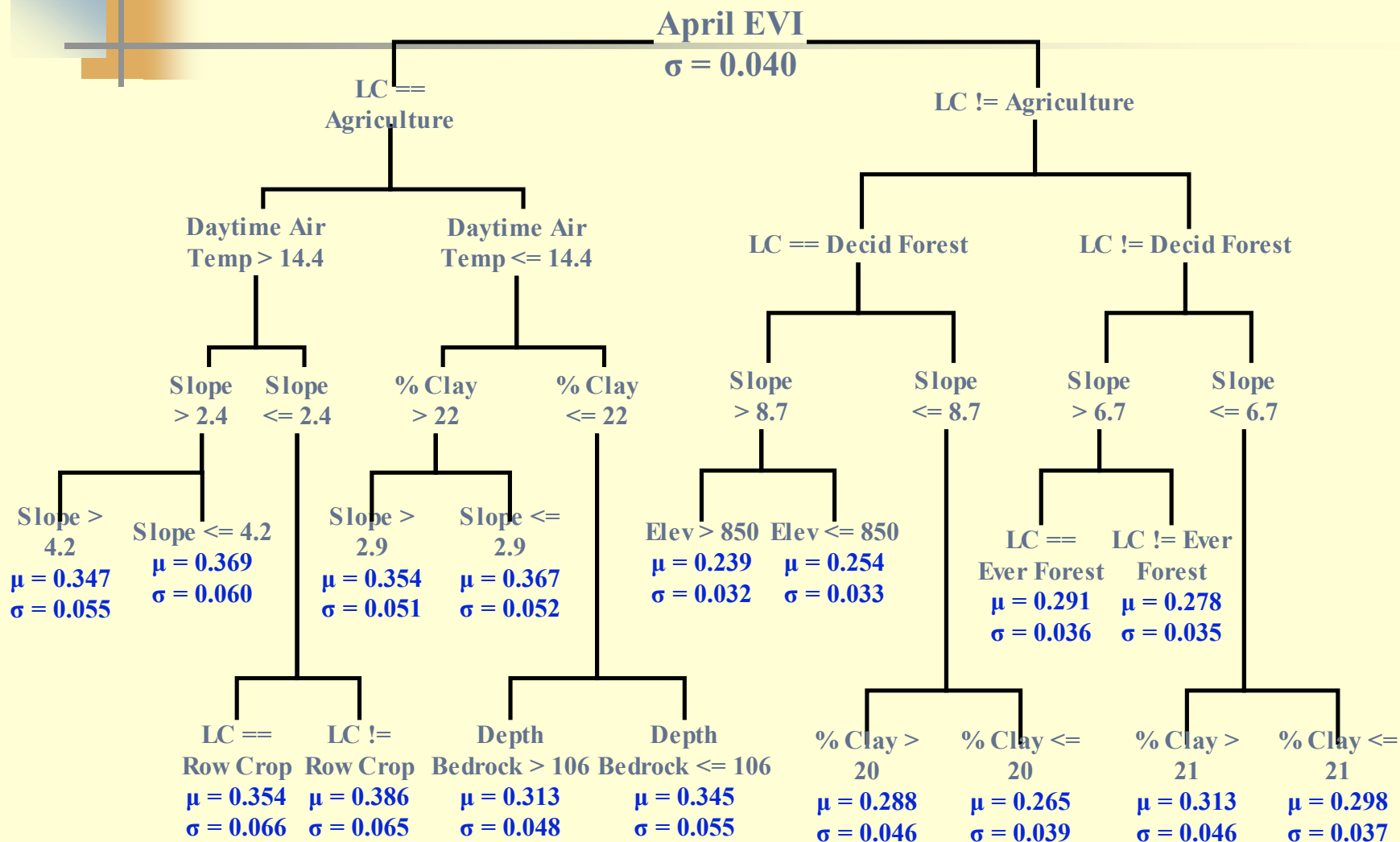


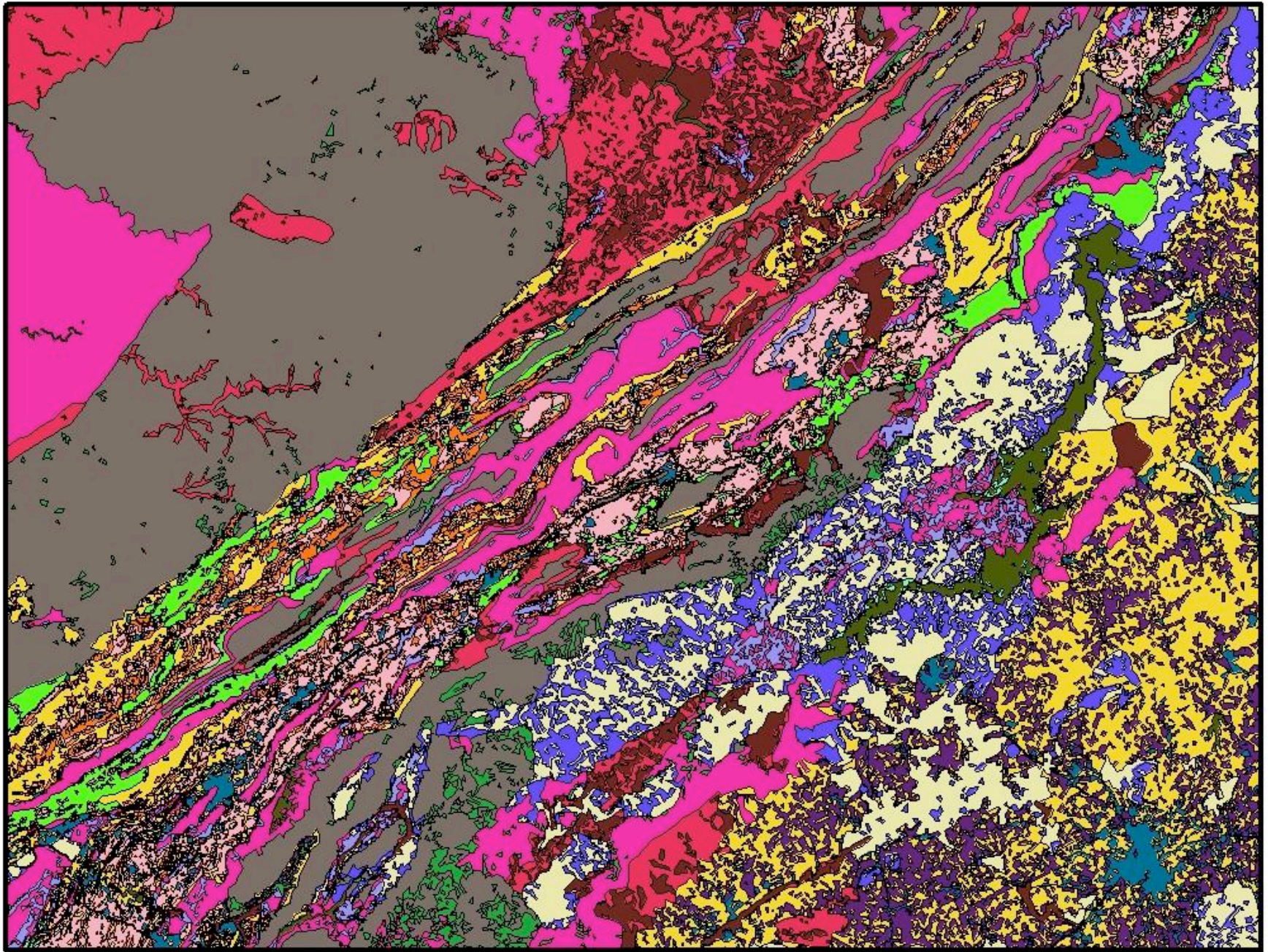
Two-Meter Daytime Temperature (°C)

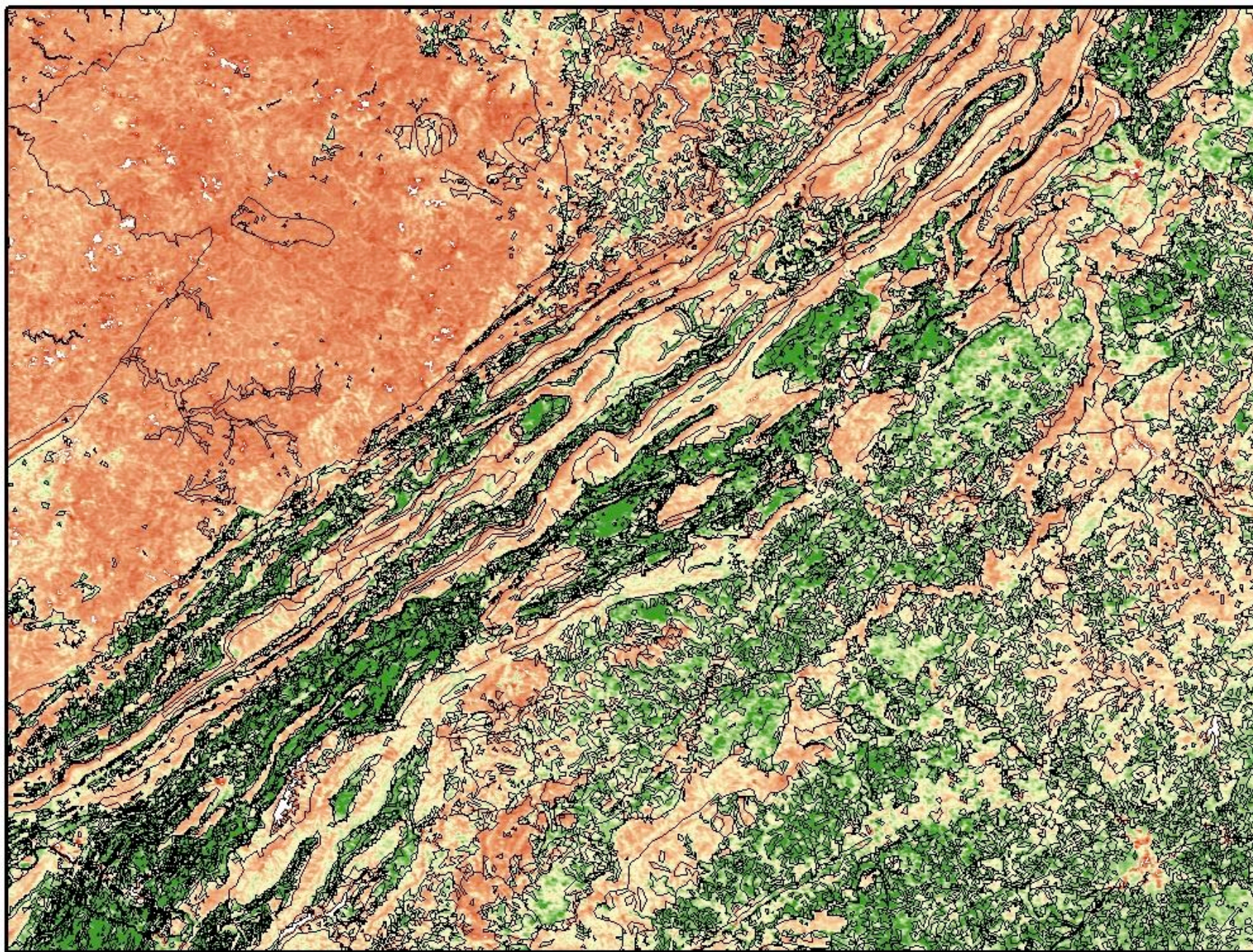


Regression Tree

2000-2004 Average
2/3 Training
1/3 Prediction
Bootstrapping







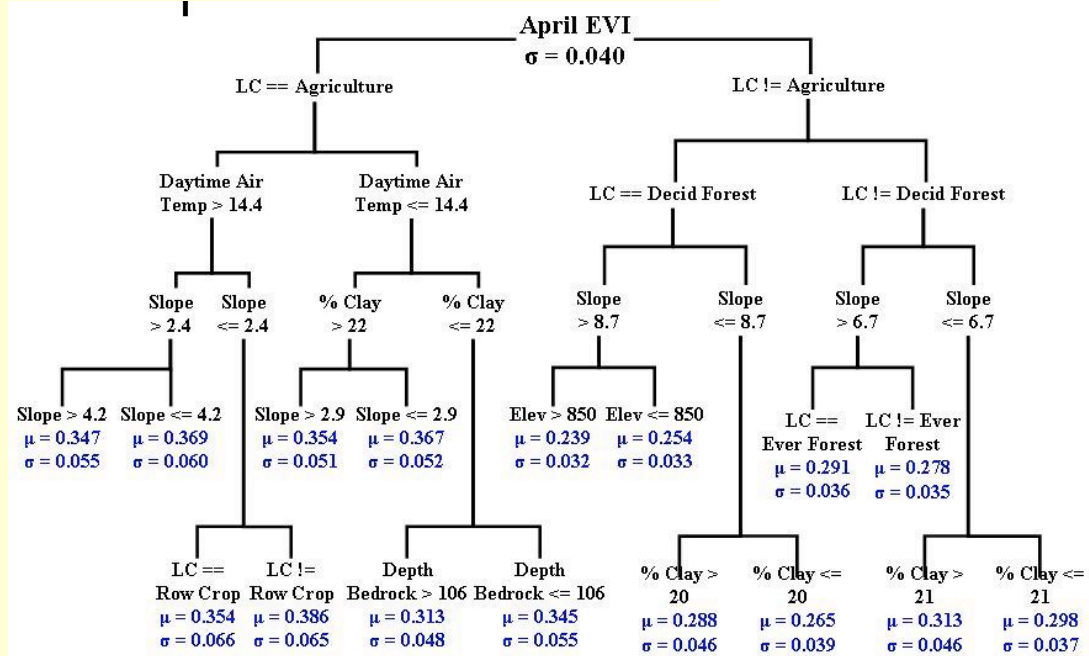
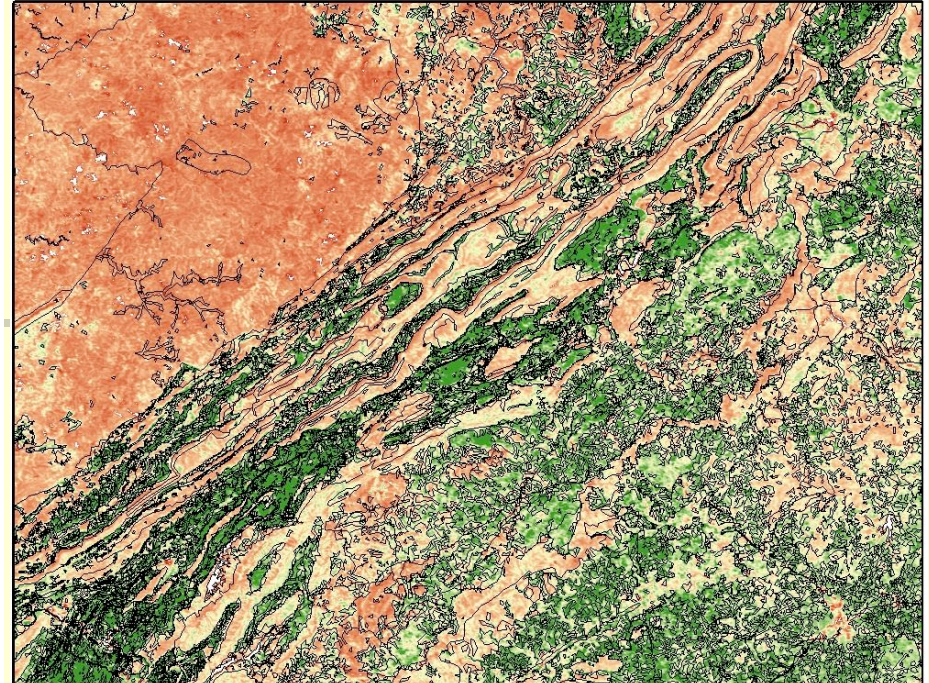
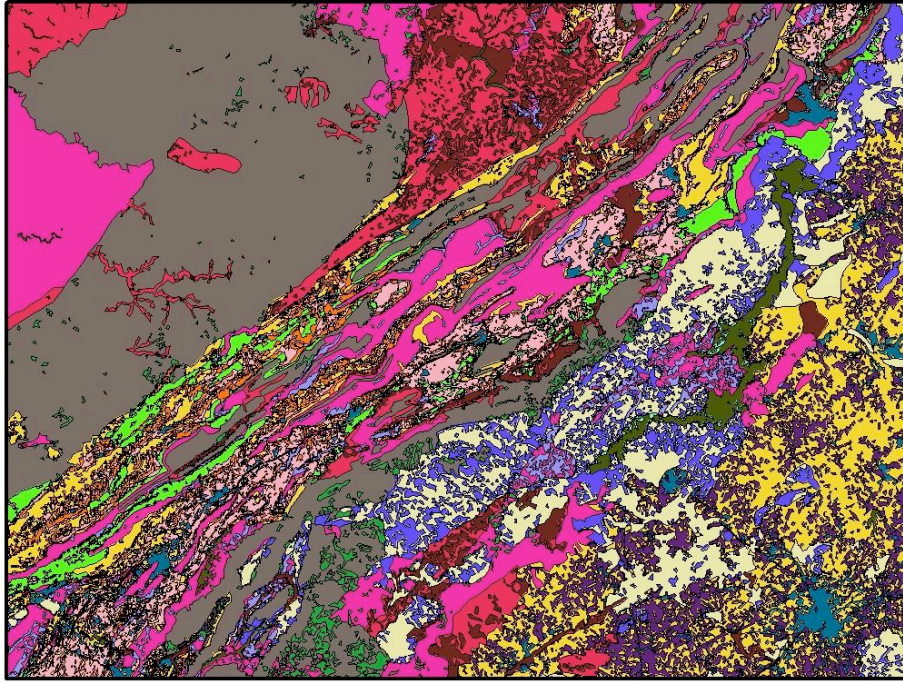
EVI

0.11

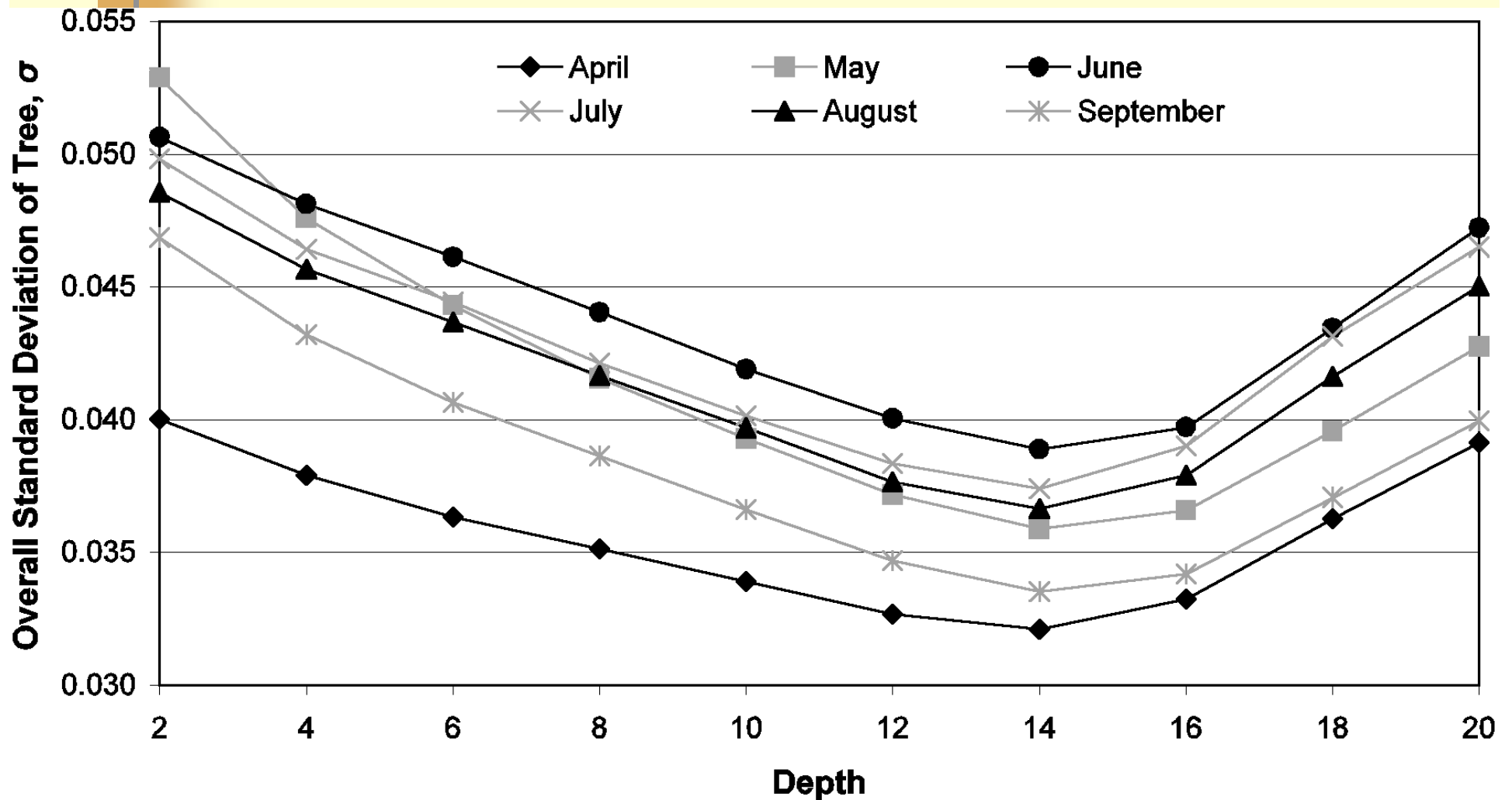
0.31

0.51

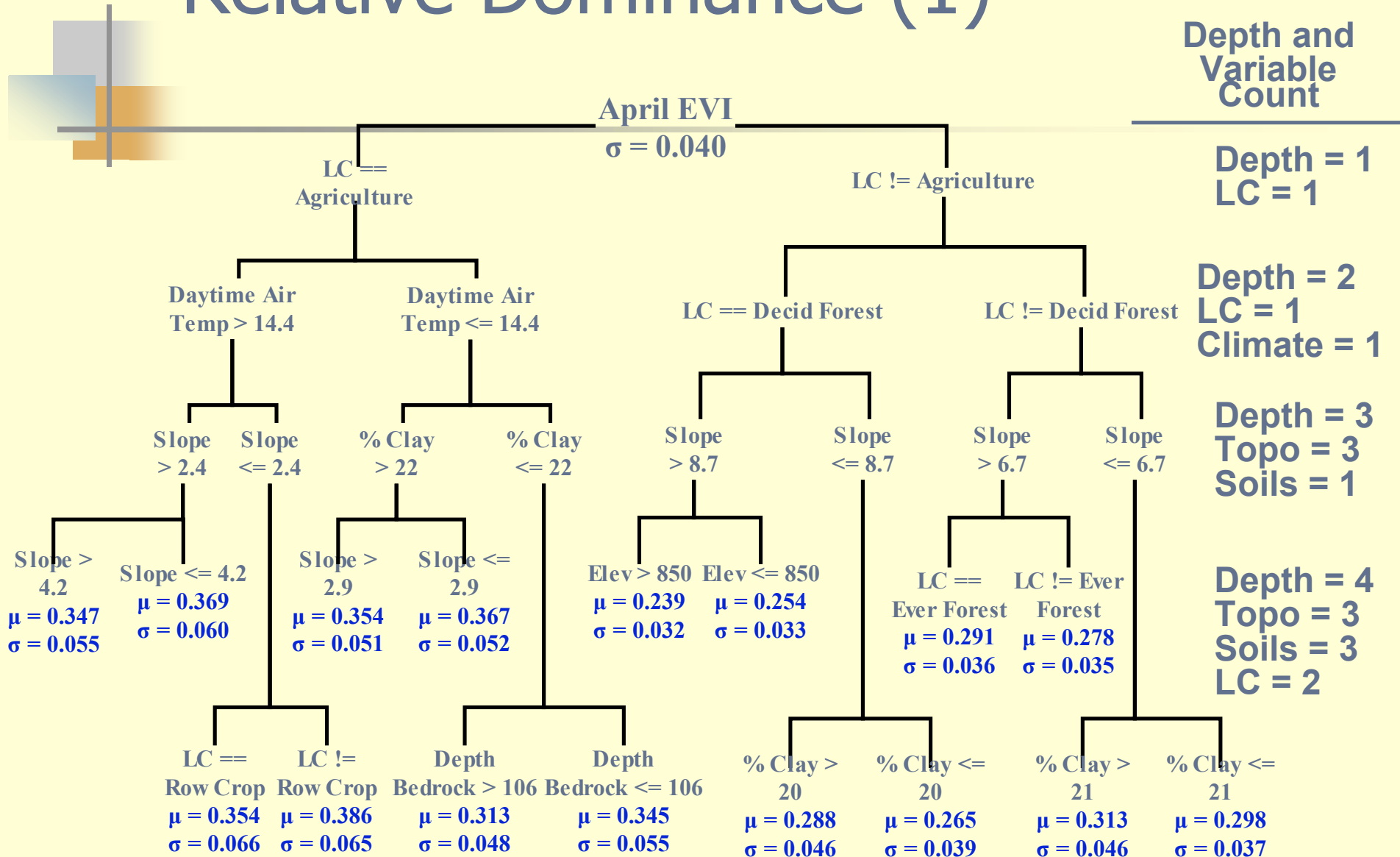
0.71



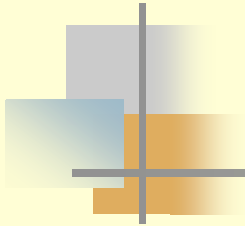
Tree Depth vs. Error



Relative Dominance (1)



Relative Dominance (2)



$$R_V = \sum_{i=1}^D \frac{n}{i}$$

V = variable

= {topography, soils, land cover, climate}

D = depth of tree

n = number of occurrences of variable

Example with $D = 4$

$$R_{\text{topography}} = \frac{0}{1} + \frac{0}{2} + \frac{3}{3} + \frac{3}{4} = 1.750$$

$$R_{\text{soils}} = \frac{0}{1} + \frac{0}{2} + \frac{1}{3} + \frac{3}{4} = 1.083$$

$$R_{\text{land cover}} = \frac{1}{1} + \frac{1}{2} + \frac{0}{3} + \frac{2}{4} = 2.000$$

$$R_{\text{climate}} = \frac{0}{1} + \frac{1}{2} + \frac{0}{3} + \frac{0}{4} = 0.500$$

$$\text{Sum} = 5.333$$

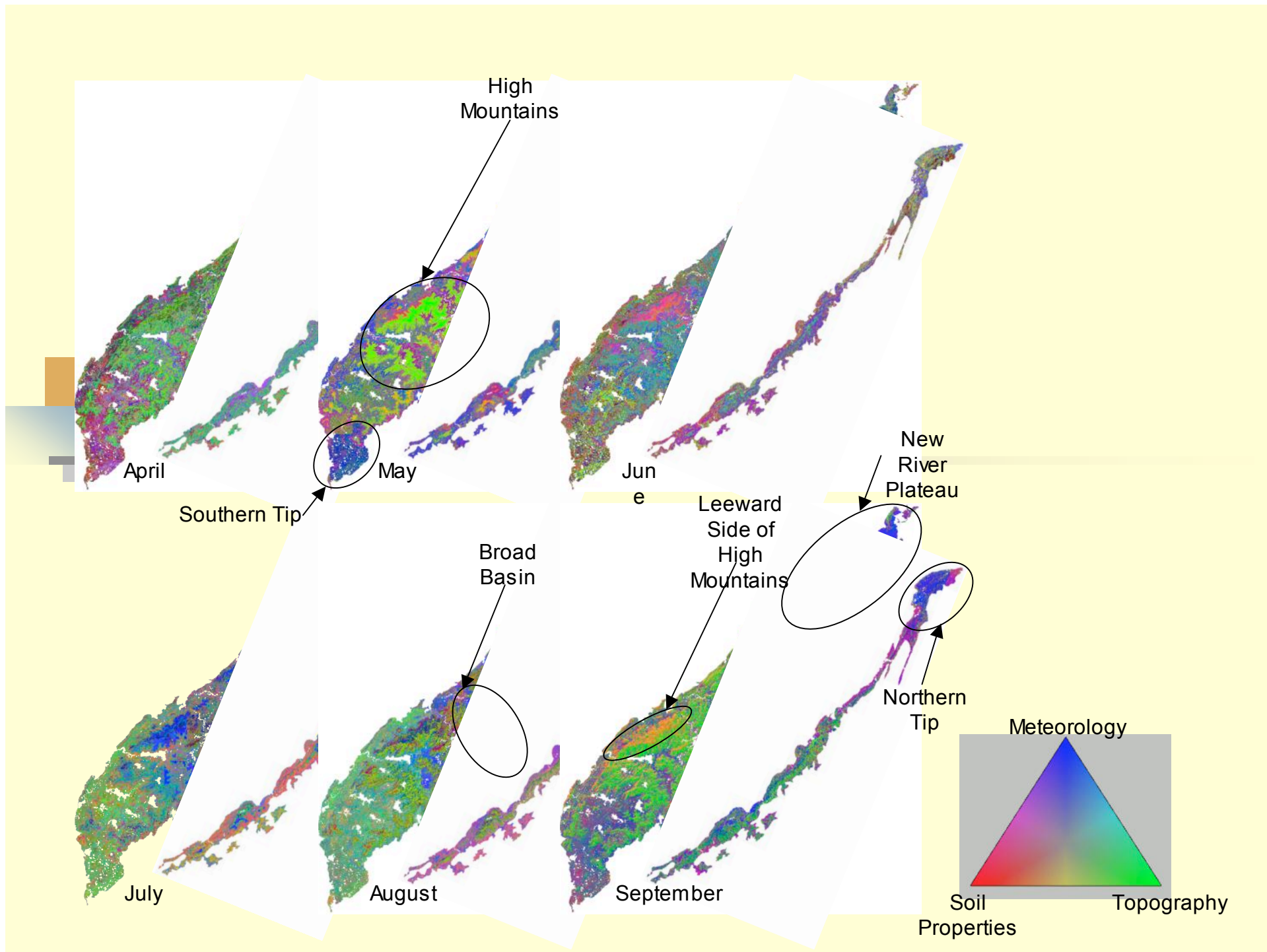
$$R_{\text{topography}} = \frac{1.750}{5.333} = 0.3281 = 32.81\%$$

$$R_{\text{soils}} = \frac{1.083}{5.333} = 0.2031 = 20.31\%$$

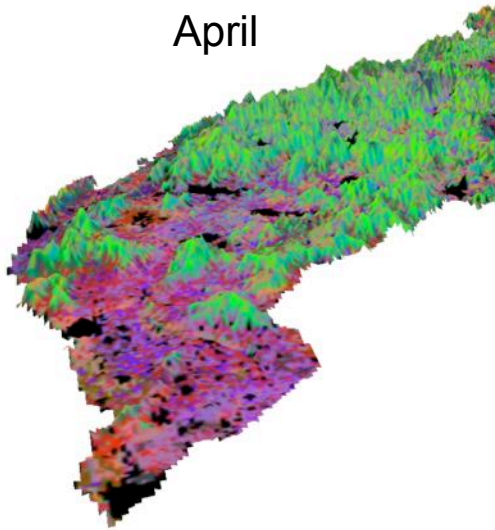
$$R_{\text{land cover}} = \frac{2.000}{5.333} = 0.3750 = 37.50\%$$

$$R_{\text{climate}} = \frac{0.500}{5.333} = 0.0938 = 9.38\%$$

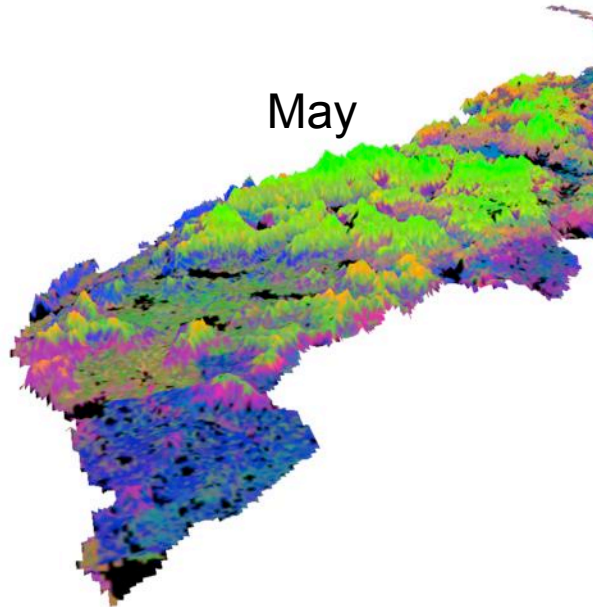
$$\text{Sum} = 100\%$$



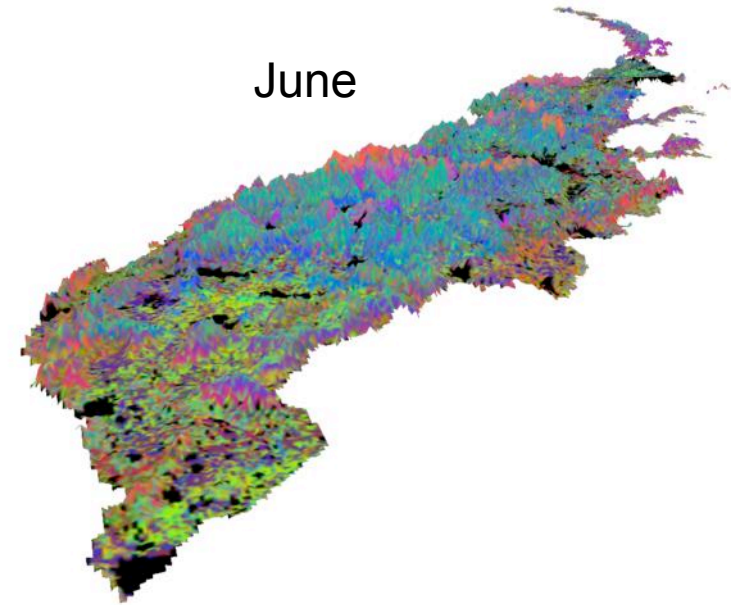
April



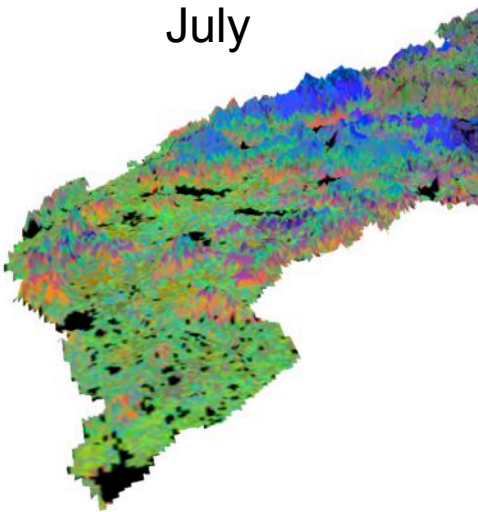
May



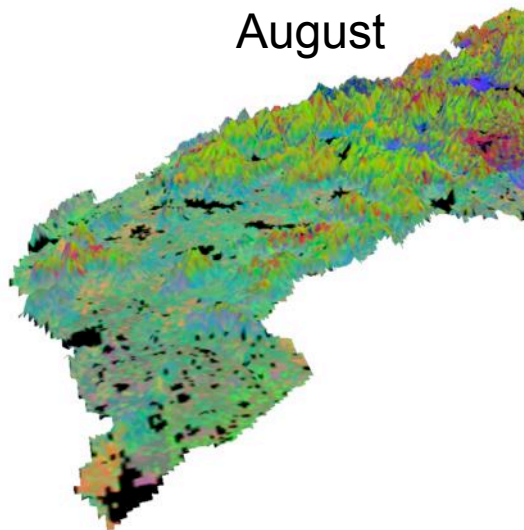
June



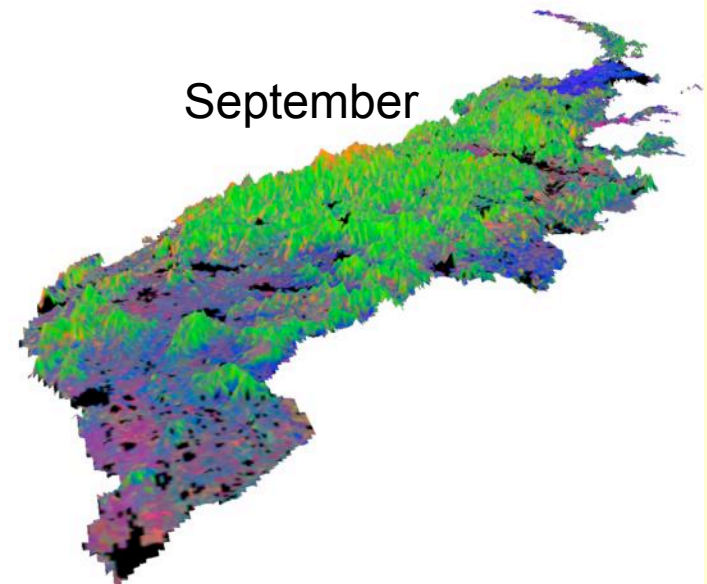
July



August

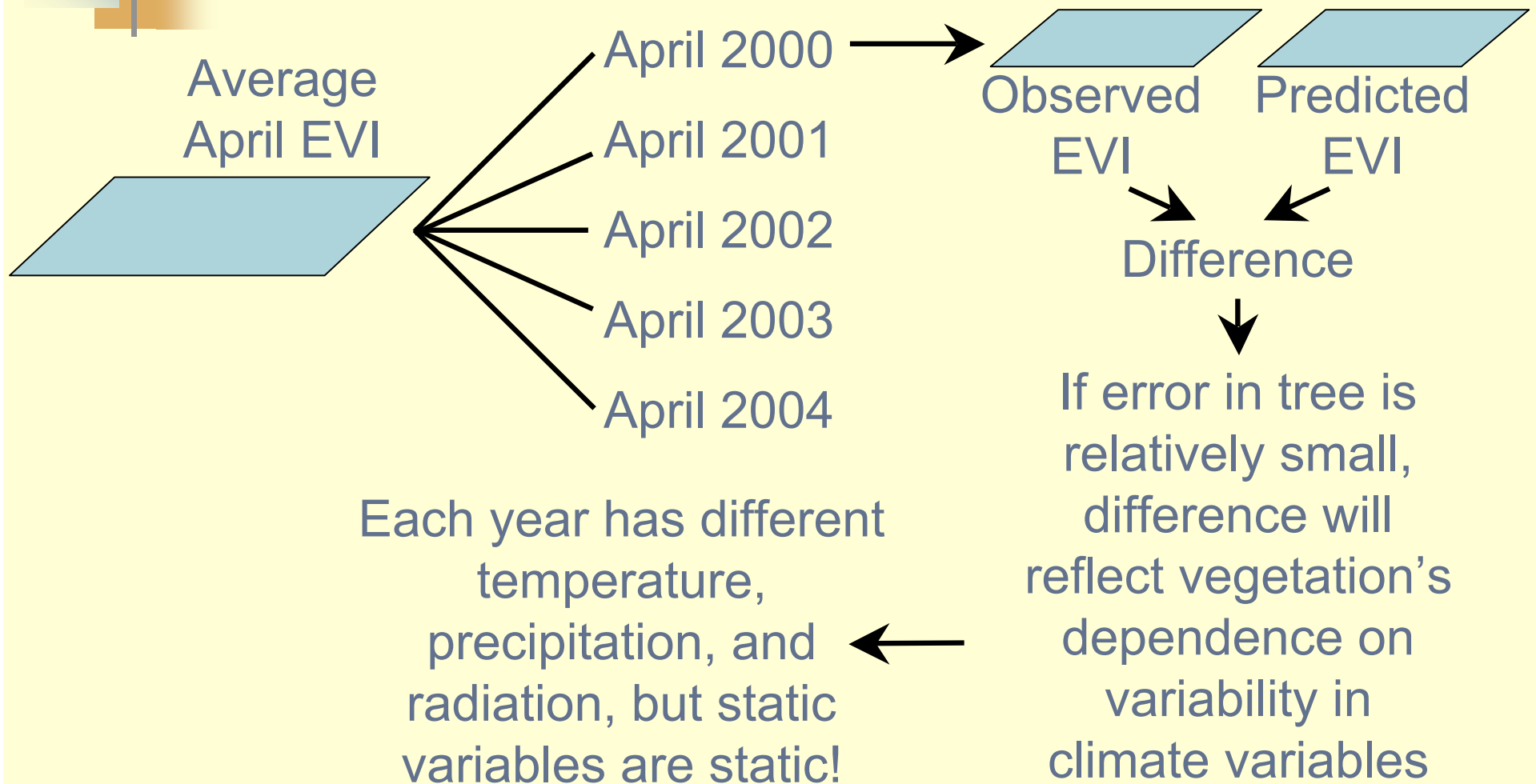


September

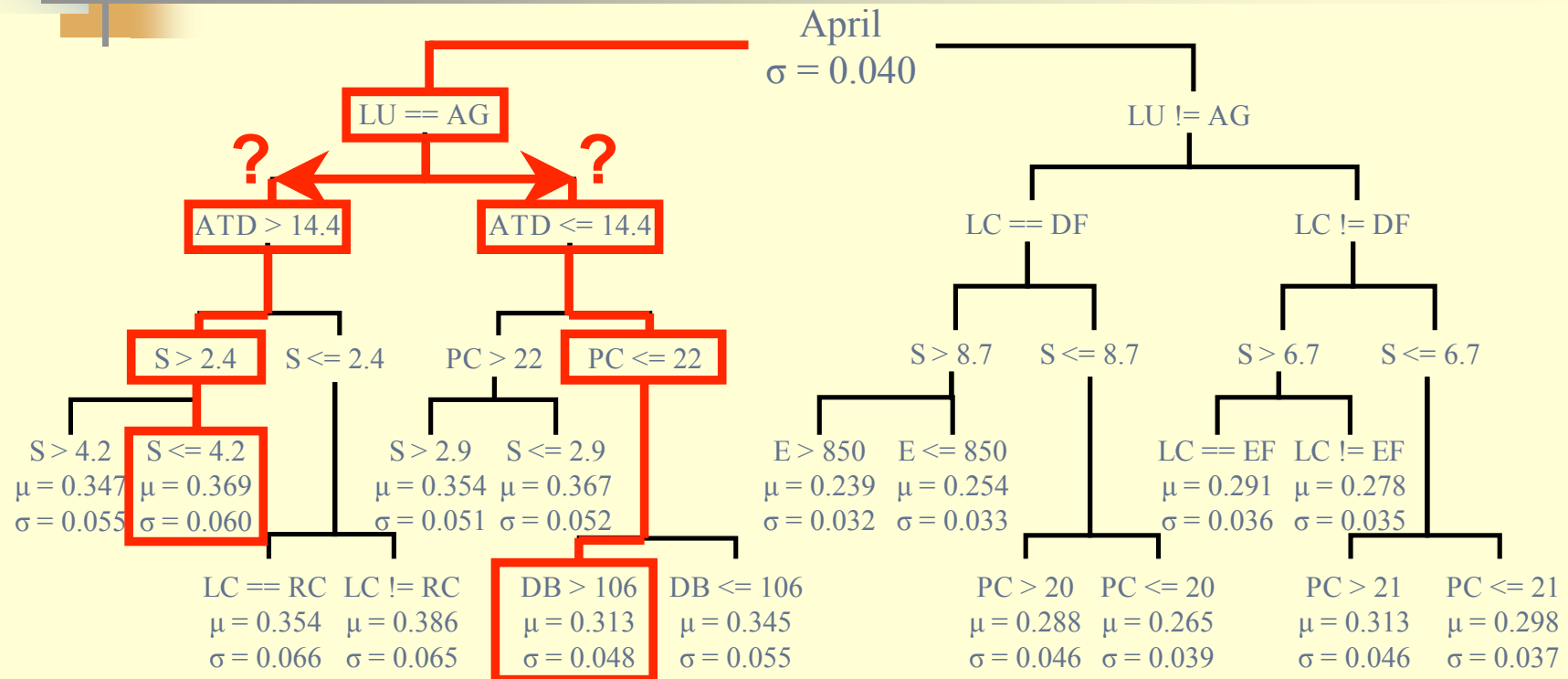


raphy

How Can We Use These Trees for Prediction?



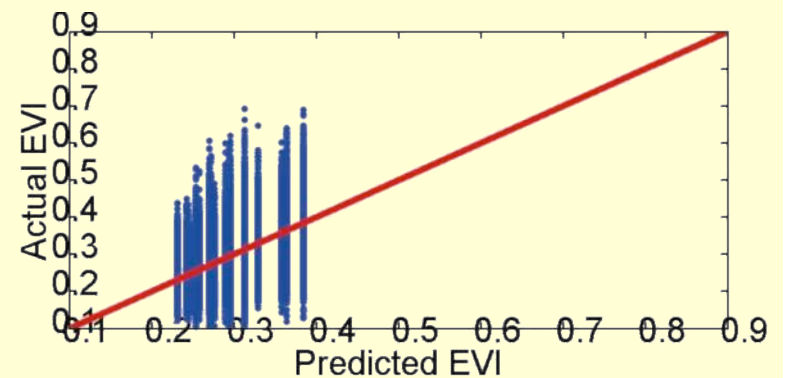
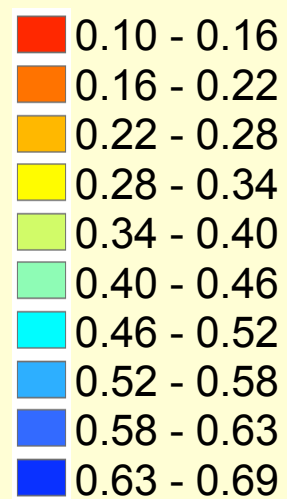
Regression Tree Prediction Model



April 2004

Actual EVI

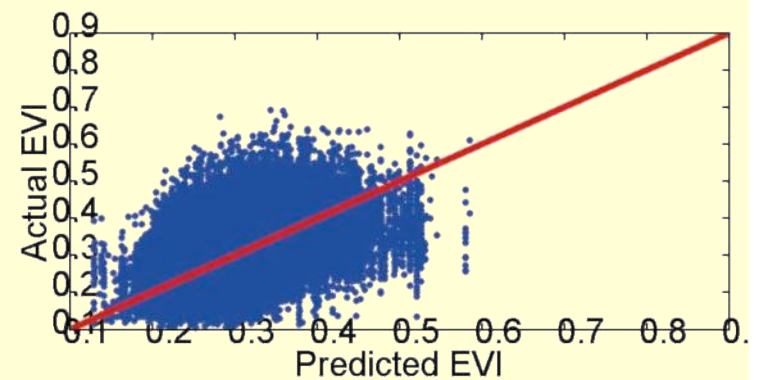
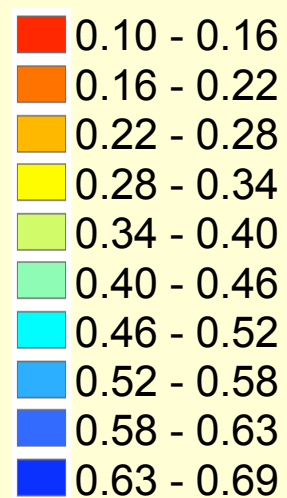
Predicted
Level 4



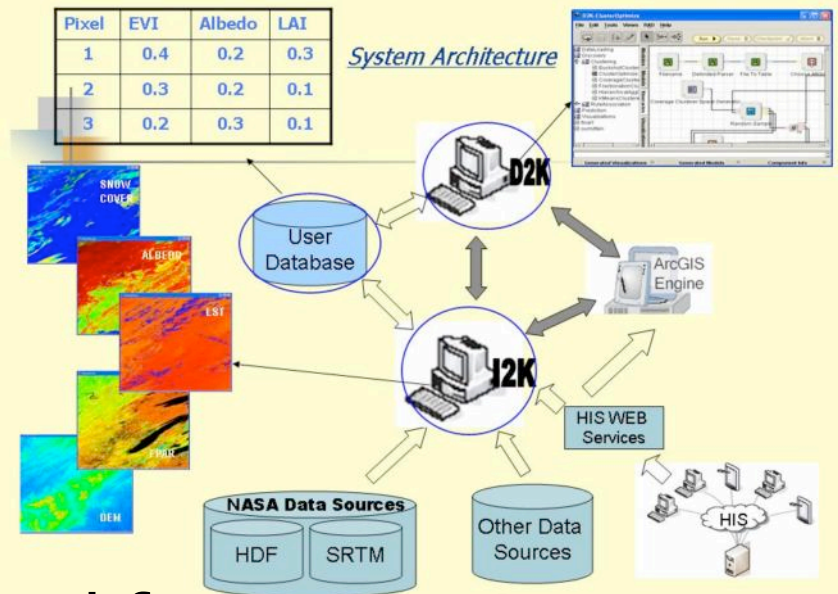
April 2004

Actual EVI

Predicted
Level 14



Summary



- Architecture is being developed for integration remote sensing and GIS data for data mining applications
- Initial results suggest the ability to perform detailed (pixel-by-pixel) investigations in space-time domain
 - Dependencies over multitude of variables can be analyzed
- Next steps involves expanding the ability for analyses over much larger domain sizes